DISSERTATION

ERK signal duration decoding by mRNA dynamics

Florian Uhlitz

BLÜTHGEN GROUP
COMPUTATIONAL
MODELLING IN
MEDICINE

ERK signal duration decoding by mRNA dynamics

D I S S E R T A T I O N

zur Erlangung des akademischen Grades

doctor rerum naturalium
(Dr. rer. nat.)
im Fach Biologie

eingereicht an der
Lebenswissenschaftlichen Fakultät
der Humboldt-Universität zu Berlin

von
M.Sc. Florian Sören Uhlitz

Präsidentin der Humboldt-Universität zu Berlin:
Prof. Dr.-Ing. Dr. Sabine Kunst

Dekan der Lebenswissenschaftlichen Fakultät:
Prof. Dr. Bernhard Grimm

Gutachter:
1. Prof. Dr. Christine Sers
2. Prof. Dr. Markus Landthaler
3. Prof. Dr. Nils Blüthgen

Eingereicht am: 29.11.2018
Tag der mündlichen Prüfung: 22.03.2019

Contents

Zusammenfassung

Der RAF-MEK-ERK-Signalweg steuert grundlegende, oftmals entgegengesetzte zelluläre Prozesse wie die Proliferation und Apoptose von Zellen. Die Dauer des vermittelten Signals wurde als entscheidener Faktor für die Steuerung dieser Prozesse identifiziert. Es ist jedoch nicht eindeutig geklärt, wie die verschiedenen früh und spät reagierenden Genexpressionsmodule kurze und lange Signale unterscheiden können und durch welche kinetischen Merkmale ihre Antwortzeit bestimmt wird. In der vorliegenden Arbeit wurden sowohl Proteinphosphorylierungsdaten als auch Genexpressionsdaten aus HEK293-Zellen gewonnen, die ein induzierbares Konstrukt des Proto-Onkogens RAF tragen. Hierbei wurde ein neues Genexpressionsmodul identifiziert, dass sich aus sofort induzierten aber spät antwortenden Genen zusammensetzt. Es unterscheidet sich in der Genexpressionsdynamik und Genfunktion von anderen Modulen, und wurde mit Hilfe mathematischer Modellierung experimenteller Daten identifiziert. Es wurde festgestellt, dass diese Gene aufgrund von langen Halbwertszeiten der vermitteltenden mRNA in der Lage sind spät auf das eingehende Signal zu reagieren und die Dauer des Signals in die Amplitude der Genantwort zu übersetzen. Trotz der langsamen Akkumulation und damit späten Antwortzeit, konnte aufgrund einer GC-reichen Promoterstruktur zunächst vermutet und mit Hilfe eines Markerverfahrens bestätigt werden, dass die Transkription dieser Gene instantan mit Beginn der ERK-Aktivierung startet. Eine vergleichende Analyse zeigte, dass das Prinzip der Signaldauer-Entschlüsselung in PC12-Zellen und MCF7-Zellen, zwei paradigmatischen Zellsystemen für die ERK-Signaldauer, konserviert ist. Insgesamt deuten die Ergebnisse der Untersuchung darauf hin, dass das neu identifizierte Genexpressionsmodul der Entschlüsselung der ERK-Signaldauer dient und das mRNA Halbwertszeiten sowohl hierfür, als auch für die zeitliche Abfolge der Genantwort eine entscheidende Rolle spielen.

Abstract

The RAF-MEK-ERK signalling pathway controls fundamental, often opposing cellular processes such as proliferation and apoptosis. Signal duration has been identified to play a decisive role in these cell fate decisions. However, it remains unclear how the different early and late responding gene expression modules can discriminate short and long signals and what features govern their timing. Both protein phosphorylation and gene expression time course data was obtained from HEK293 cells carrying an inducible construct of the proto-oncogene RAF. A new gene expression module of immediate-late genes (ILGs) distinct in gene expression dynamics and function was identified by mathematical modelling. It was found that mRNA longevity enables these ILGs to respond late and thus translate ERK signal duration into response amplitude. Despite their late response, their GC-rich promoter structure suggested and metabolic labelling with 4SU confirmed that transcription of ILGs is induced immediately. A comparative analysis showed that the principle of duration decoding is conserved in PC12 cells and MCF7 cells, two paradigm cell systems for ERK signal duration. Altogether, the findings of this study indicate that ILGs decode ERK signal duration and that both decoding capacity and gene expression timing are governed by mRNA half-life.

Opening remark

Research is a joint effort. Any new insight is enabled by past findings of fellow scientists, and collaborations between researchers are essential to integrate knowledge from different fields. Along these lines, the findings presented in my thesis are based on a collective effort that resulted in a manuscript published in Molecular Systems Biology (Uhlitz et al. 2017). I was the sole first author of this publication, wrote the manuscript, performed all computational analyses and mathematical modelling, was involved in experimental design and conducted part of the experiments, namely cell culture, RT-qPCR experiments, RNA-sequencing library preparation, live-cell imaging and flow cytometry experiments. However, this work would not have been possible without the remarkable contributions of my collaborators. Anja Sieber, Raphaela Fritsche-Günther and Nadine Lehmann performed all microarray experiments, western blots and phosphoplex assays. Anja Sieber and Emanuel Wyler conducted 4SU-labelling experiments. Bertram Klinger and Nils Blüthgen designed microarray experiments. Nils Blüthgen supervised the project. Johannes Meisig and Markus Landthaler gave feedback on the manuscript. The manuscript was published under a Creative Commons license (CC BY 4.0) allowing for free republication. Whereas many paragraphs and figures in this thesis were directly adopted from the manuscript, many more were extensively edited or newly added.

1 General introduction

Wie wir uns auch drehen und wenden, wir kommen zuletzt auf die Zelle zurück.
—Rudolf Virchow (1859, p. 15)

The human body roughly consists of about ten trillion (10^{13}) human cells (Bianconi et al. 2013), each of which consists of the same genomic blueprint. Nonetheless, they do not build a shapeless lump of cells, but form and maintain a complex organism with highly specialised tissues and organs. The large variety of cells can be explained by the fact that different cells make different use of the genetic information they are equipped with, i.e. they only express cell type–specific subsets of genes.

These subsets of expressed genes eventually determine a cell's form and function, its phenotype. However, the composition and amount of expressed genes can be modulated by effector molecules like hormones, growth factors or cytokines secreted by other cells in the body. Cellular receptors recognize these effector molecules which in turn elicit the activation of signalling networks inside the cell. In a process called signal transduction the perceived stimulus is propagated to the cellular nucleus where transcriptional activators and repressors regulate gene expression. Depending on the identity and the concentration of the effector molecule, as well as on the resulting strength and duration of the intracellular signal, the gene regulatory changes ultimately lead to a defined cellular response.

Cellular responses can be subtle, like the release of additional effector molecules to propagate signals to neighbouring cells. Yet, other responses may entirely redefine a cell's form and function. For example, certain growth factor treatments may lead to cellular differentiation, a process in which cells change from one cell type to another. Considering different growth factor treatments in particular, the importance of signal duration for cellular responses becomes likewise apparent. In response to short growth factor–induced signalling pulses, cells commonly undergo cell division. On the contrary, an extensive long-lasting signal can drive them into cell death, permanently transform them or even turn them into cancer cells.

But how do signalling dynamics translate to cellular phenotypes in the first place? How can different signal durations alter the composition of responding gene sets, which in turn determine cellular fates?

In my thesis, I combine experimental and computational methods to provide new insights on these questions. In short, I present evidence that gene product stability, i.e. the lifetime of biological molecules, is at the heart of these processes, governing a cell's capability to translate signalling dynamics to cellular outcome. In the first part (ch. 2), I approach the subject from a theoretical point of view, introducing a simple mathematical model of gene expression and its implications for signal duration decoding. Based on computational simulations I arrive at the hypothesis that gene product stability shapes response timing and allows for signal duration decoding. Subsequently, the second part of my thesis (ch. 3) introduces a newly identified temporal cluster of genes (immediate-late genes, ILGs) which are induced immediately but respond late due to high gene product stability. In the last part (ch. 4), I demonstrate that these characteristics potentially enable ILGs to mediate between signalling dynamics and cellular outcome and that this principle is conserved across different experimental model systems. A more detailed outline of my thesis will be provided at the end of the introduction.

On the following pages, I will first introduce important concepts underlying the findings presented in my thesis. To begin with, I will touch upon the general concept of cell signalling, introduce the signalling network of interest, the extracellular signal regulated kinase (ERK) signalling network, and discuss what is currently known about the gene regulatory response to ERK signalling. To motivate the subject of my thesis, I will then outline its important role in cancer and how novel cancer therapies sparked from cell signalling research. Lastly, I will provide some information on encoding and decoding of cell signalling dynamics with an emphasis on what is known about signal duration encoding and decoding and discuss how new insights in this field will help to better understand common principles of cell signalling.

1.1 ERK signalling and its gene regulatory response

Cells constantly adjust their behaviour to changes in their environment. They do so by translating extracellular signals into cellular responses like cell division, changes in metabolism, cell shape or cell identity. This capability is not limited to multicellular organisms. Unicellular organisms like bacteria and yeast use cell signalling to respond to changes in availability of nutrients or to coordinate their behaviour with conspecifics, including their motility or mating. In multicellular organisms cell signalling is commonly induced by different kinds of cell-to-cell communication. Cells can stimulate or inhibit other cells in a contact-dependent manner that requires direct interaction of membrane-bound signalling molecules and target cell receptors, or in a contact-independent manner by secreting signalling molecules into their surroundings. When distributed via the bloodstream, secreted signalling molecules can act over long distances (endocrine signalling), which is the case for hormones. Whereas hor-

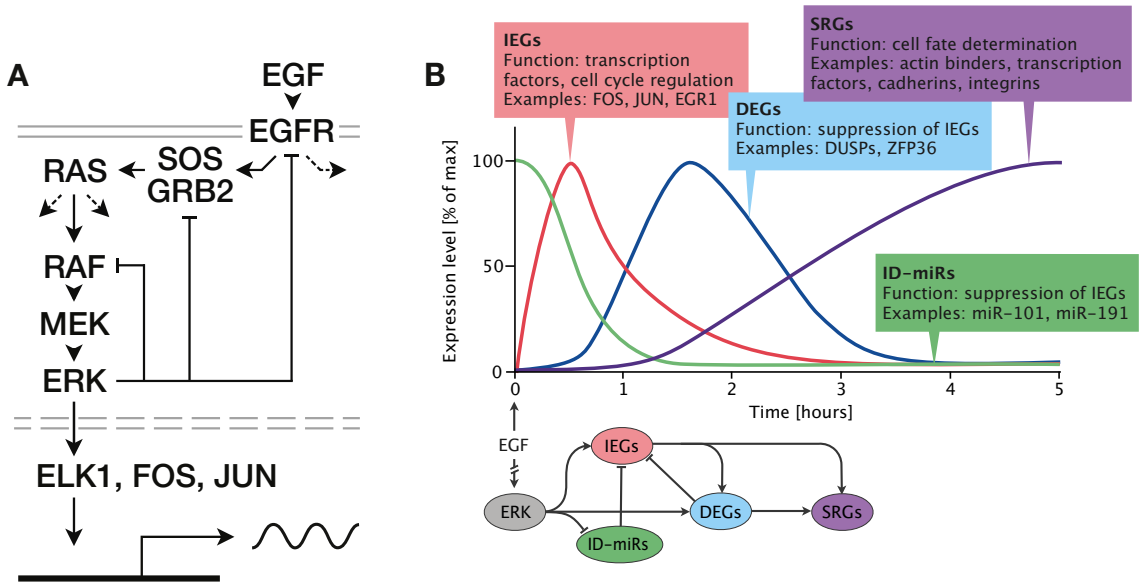


Figure 1.1 ERK signalling network and its gene regulatory response.
A: The ERK signalling network transduces signals perceived at the cellular membrane to the nucleus, eliciting a complex gene regulatory response. For a description of the cascade, refer to the main text. Arrow heads indicate activation and orthogonal line ends indicate inhibition. Dashed arrows indicate major branching points in the signalling network. Grey solid lines indicate cellular membrane and grey dashed lines indicate nuclear membrane. The bottom element indicates an active promoter transcribing RNA. **B**: Upper panel: The gene regulatory response to EGF-mediated ERK signalling is composed of different waves of gene sets. These include primary response gene sets of immediate-early genes (IEGs, red), delayed-early genes (DEGs, blue), as well as secondary response genes (SRGs, violet) and immediately downregulated microRNAs (ID-miRs, green). Lower panel: The responding modules form an interconnected gene regulatory network. Divided arrow indicates that intermediate steps from EGF to ERK are not shown. Figure B was adapted from Avraham and Yarden (2011).

mones can affect the entire organism, growth factors commonly act as local mediators. They are sensed by neighbouring cells (paracrine signalling) or by the secreting cells themselves (autocrine signalling). Mechanistically, specialised cell-surface receptors or intracellular receptors recognize the molecular structure of a particular signalling molecule (key-and-lock model). The perceived signal is then transmitted through a cell-specific network of intracellular signalling proteins which regulate the activity of cell fate governing effector proteins like metabolic enzymes, transcription factors or cytoskeletal proteins. It is important to stress that the diversity of cellular responses is primarily enabled by the complex integration of a rather small set of signalling networks which are able to encode different signalling inputs in a large variety of signalling dynamics.[1]

[1]For a comprehensive primer on cell signalling, refer to Alberts et al. (2014, pp. 813-888).

One of the most extensively studied signalling networks is the extracellular signal–regulated kinase (ERK) signalling network (Davis 1993), as it controls a wide range of both normal and aberrant cellular phenotypes. On the one hand, it regulates crucial cellular responses including cellular survival, growth, proliferation, differentiation, migration and apoptosis (Oda et al. 2005). On the other hand, elements of the network are often involved in cancer initiation and progression (Roberts and Der 2007). In general, it belongs to the family of MAP kinase signalling networks and its network topology can vary across cell types or in response to different signalling molecules, resulting in a variety of different signalling dynamics like transient, sustained or oscillatory activity patterns.[2] Commonly, ERK signalling involves a cascade of signalling events starting at epidermal growth factor receptor (EGFR) and culminating with the activation of ERK (fig. 1.1). More precisely, EGFR proteins sense epidermal growth factor (EGF) and form catalytically active dimers. Dimerised, they activate guanine nucleotide exchange factor SOS (son of sevenless) with help of adaptor molecule GRB2 (growth factor receptor-bound protein 2). Activated SOS/GRB2 complex loads members of the small GTPase protein family RAS with GTP. Loaded with GTP, RAS proteins recruit kinases of the RAF protein family to the membrane. At the membrane, RAF proteins are activated by phosphorylation. The cascade continues with RAF proteins phosphorylating MEK proteins MEK1 and MEK2 and MEK proteins eventually phosphorylating the terminal kinases ERK1 and ERK2.[3]

Along the way from EGFR to ERK many other signalling elements can get activated or inactivated at different branching points, emphasising the interconnected nature of signalling networks in general. Pathway divergence starts at the level of EGFR, as it can be recognised by other adaptor proteins as well, including Src homology 2 domain containing transforming protein (SHC) and GRB2-associated binding protein 1 (GAB1) (Oda et al. 2005). Although both proteins share RAS as a target with GRB2, they can potentially activate other signalling branches like phosphatidylinositol 3-kinase (PI3K) and Src signalling. GRB2 serves as the next branching node in the network. In addition to RAS activation it can, for instance, mediate EGFR ubiquitination by recruiting an ubiquitin ligase (CBL) resulting in EGFR degradation (Avraham and Yarden 2011). Above all, RAS might be considered as the major branching point along the signalling axis from EGFR to ERK. Next to activation of RAF isoforms BRAF and RAF1, it can serve as an additional link to PI3K signalling. In fact, ERK signalling and PI3K signalling are often studied mutually, as they cross-talk extensively and can regulate each other both positively and negatively (Mendoza, Er, and Blenis 2011; Stelniec-Klotz et al. 2012). On top of PI3K regulation, RAS can mediate signalling via T

[2] Other classical MAP kinase signalling networks are the p38 MAP kinase and the JNK signalling networks. In this thesis however, the term MAP kinase signalling refers to ERK signalling unless stated otherwise.
[3] In the following, MEK1 and MEK2 are referred to as MEK, and ERK1 and ERK2 are referred to as ERK.

lymphoma invasion and metastasis protein 1 (TIAM1) or Ral guanine nucleotide dissociation stimulator (RalGDS). These and additional RAS targets which often depend on cell type and stimulus are reviewed in Rajalingam et al. (2007) and Buday and Downward (2008). Whereas almost no other *bonda fide* targets are known for RAF and MEK (Matallanas et al. 2011), a large compendium of targets can potentially be regulated by ERK (Ünal, Uhlitz, and Blüthgen 2017). Most prominently, a defined subset of transcription factors is activated by nuclear translocated ERK to orchestrate gene regulatory changes in response to sensed growth factor signals. Among others, these commonly include ELK1 and AP-1 complex elements FOS and JUN (Avraham and Yarden 2011), but the exact composition and the resulting transcriptional response depends on different properties such as cell type identity and signal duration and will be discussed further below. At the same time, many cytosolic ERK targets can be regulated as well. For example, ERK can phosphorylate EGFR (Northwood et al. 1991) and thereby impair cross activation of receptor dimers (Sato et al. 2013). ERK-mediated disruption of SOS/GRB2 complex and RAS/RAF binding (Ueki et al. 1994; Corbalan-Garcia et al. 1996) further stresses the importance of feedback mechanisms as part of the ERK signalling network (reviewed in Lake, Corrêa, and Müller 2016). Together with transcriptionally induced negative regulators like Sprouty (SPRY) proteins and dual-specificity phosphatases (DUSPs) (reviewed in Mason et al. 2006; Kidger and Keyse 2016), these feedback elements confer robustness to the ERK signalling network and allow for adaptation to a broad range of signalling inputs (Fritsche-Guenther et al. 2011). Apart from its major role as a kinase, ERK has also been reported to function independent of its kinase domain. For example, it can bind to lamin A to co-localise with nuclear substrates like FOS which can likewise bind to this component of the nuclear envelope (Worman and Bonne 2007). Without any doubt, the most peculiar kinase-independent function of ERK is its reported ability to directly act as a transcriptional repressor by site-specific binding to DNA sequence motif G/CAAG/C (Hu et al. 2009). When binding to this motif, ERK outcompetes transcription factor C/EBPβ and blocks transcription of certain IFNγ-induced genes and is only released shortly after IFNγ is sensed by the cell. Interestingly, ERK rebinds promoters after prolonged exposure to IFNγ, concomitant with repression of IFNγ-induced genes. Accordingly, one could speculate that signal duration might not only effect kinase-dependent functions of ERK but its kinase-independent functions as well (Fowler, Sen, and Roy 2011).

Thus far, no other studies reported direct regulation of gene expression via ERK DNA binding. It is hence reasonable to assume that propagation of signalling dynamics to gene expression dynamics mainly occurs via ERK-regulated transcription factors. Yet, gene regulatory dynamics are multi-layered. In addition to gene specific transcription factors like

ELK1 or FOS, the eukaryotic transcriptional machinery consists of general initiation factors and RNA polymerase (Roeder 2003). Chromatin remodelling proteins and cell-specific co-factors add additional layers of complexity (Roeder 2003) and may all be subject to ERK regulation. Besides transcription, mRNA levels are also affected by processing rates of pre-mRNA to mRNA, including the processes of 5'UTR capping, 3'UTR polyadenylation and splicing (Moore and Proudfoot 2009). Last but not least, RNA binding proteins (RBPs) and microRNAs or other non-coding RNAs (ncRNAs) can bind mRNA to alter mRNA transcript stability which likewise affects mRNA expression levels (Brennan and Steitz 2001; Fabian and Sonenberg 2010; Avraham et al. 2010; Aitken et al. 2015). To meet this complexity, the following paragraphs will discuss what is currently known about both ERK-mediated transcriptional and post-transcriptional regulation of mRNA.

Upon activation ERK dissociates from anchoring proteins in the cytoplasm and translocates to the nucleus (Murphy and Blenis 2006; Whitmarsh 2011). In the nucleus, ERK is recruited to the chromatin and co-localises with transcriptional complexes at serum response elements (SREs) to phosphorylate its substrates *in situ* (Chow and Davis 2006; Pokholok 2006; Whitmarsh 2007; Zhang et al. 2008), causing transcription of different early and late responding gene expression modules (fig. 1.1B). These encompass two different groups of primary response genes (PRGs), namely immediate-early genes (IEGs) and delayed-early genes (DEGs), as well as secondary response genes (SRGs). The presence and temporal order of most of these modules was established and confirmed in a variety of system-wide analyses of MAP kinase–mediated gene expression. Studies started off in yeast (Posas 2000; Capaldi et al. 2008) and continued in mammalian cell lines including HeLa, MCF10A (Amit et al. 2007), T98G (Tullai et al. 2007), MCF7 (Saeki et al. 2009; Shiraishi, Kimura, and Okada 2010), H1299 (Nagashima et al. 2009) and PC12 cells (Offermann et al. 2016), to name a few.

At first, IEGs are induced and peak in expression at about 15 to 45 minutes post-stimulation. DEGs are expressed in a delayed fashion and peak about 45 to 120 minutes post-stimulation (reviewed in Avraham and Yarden 2011; Healy, Khan, and Davie 2013). So far, the differences in IEG and DEG dynamics were mainly attributed to their different promoter architecture and occupancy. IEGs have been shown to possess poised promoters with GC-rich sequences (Ramirez-Carrozzi et al. 2009) and undergo continuous cycles of histone acetylation and deacetylation (Wang et al. 2009). Their permissive promoter structure furthermore includes RNA polymerase and transcription factor occupancy prior to stimulation and a leaky expression of mRNA resulting from this (Hargreaves, Horng, and Medzhitov 2009). Given this configuration, IEG promoters do not require any chromatin remodelling or protein recruitment and are immediately transcribed upon ERK activation. Removal of

negative regulators such as negative elongation factor (NELF) and immediate downregulation of repressive microRNAs (ID-miRs) and other non-coding RNAs (ncRNAs) have been described to further aid their immediate induction (Fujita, Piuz, and Schlegel 2009; Avraham et al. 2010; Aitken et al. 2015).

Very much opposed to IEGs, promoter architecture of DEGs is less permissive and induction of DEGs is not immediate but delayed. Also, whereas IEGs mainly encode for transcription factors and therefore function as feedforward elements, DEGs have been described to serve as a negative feedback module including genes which can act as suppressors of IEGs or tumour suppressor genes in general, underlining the interconnected nature of these gene expression modules (Amit et al. 2007). For instance, DEGs encode for transcriptional repressors (e.g. KLF2, KLF6, MAFF), phosphatases to inactivate ERK (e.g. DUSPs) or RNA-binding proteins (RBPs, e.g. ZFP36) which can bind to AU-rich elements (AREs) in untranslated regions (UTRs) of IEG mRNAs (Sugiura et al. 2011; Mukherjee et al. 2014). Whereas ID-miRs promote degradation of IEG mRNAs prior to ERK activation, DEGs encoding for RBPs promote their degradation after ERK activation. Hence, ID-miRs and DEGs mutually aid rapid induction and rapid attenuation of IEGs and emphasise the relevance of mRNA stability in terms of transcript dynamics. Yet, a thorough examination of mRNA stability across the different gene expression modules responding to ERK signalling is still missing, and will thus form an important aspect to be investigated in my thesis.

Despite their different gene expression dynamics, IEGs and DEGs both belong to the supergroup of primary response genes (PRGs). PRGs were first described in 1974, when it was shown that certain genes induced upon steroid receptor activation in drosophila do not require *de novo* protein biosynthesis (PBS) as they are regulated by transcription factors expressed prior to activation of the system (Ashburner et al. 1974; Yamamoto and Alberts 1976). Experimentally, their identification was achieved by application of PBS inhibitors like cycloheximide (CYHX) prior to the treatment inducing their response. This approach allowed discrimination of PRGs which do not depend on PBS and secondary response genes (SRGs) which do require the transcription and translation of PRG transcription factors. Due to their PRG dependence SRGs respond later than PRGs and commonly peak about two to ten hours post stimulation (Avraham and Yarden 2011). It has been shown that SRGs are often involved in cell fate decision processes, and accordingly their composition highly depends on cell type or stimulus (Avraham and Yarden 2011). Again, this is opposed to PRGs which are more conserved in their composition across different cell types and stimulations.

Altogether, the gene expression response to ERK signalling is a complex process which involves different modules with different dynamic properties and different molecular functions. It has been shown that these modules form a gene regulatory network including positive

feed forward and negative feedback motifs (fig. 1.1B), but it is still poorly understood how the different modules decode signalling dynamics to mediate between signalling inputs and phenotypic outcomes. In my thesis I introduce a novel gene expression module of immediate-late genes (ILGs) and I demonstrate that these genes bear an intrinsic decoding mechanism potentially allowing for a direct involvement in cell fate decisions. Further, I describe in detail how this module relates to the other previously described gene expression modules of IEGs, DEGs and SRGs. In a comparative analysis I investigate how the different modules interpret different signalling inputs and what features constitute their ability to do so. Here, an important notion will be to discriminate whether gene expression modules only relay signalling dynamics or truly decode them. But first, the next section will motivate the general need for a more thorough understanding of ERK signalling and its gene regulatory response by discussing the role of ERK signalling in cancer.

1.2 ERK signalling in cancer

In general, cell signalling is key for every organism to survive. It provides cells with the ability to adapt to an ever-changing environment and to communicate with each other. But it comes at a cost. Whenever signalling proteins are altered in their function, for example by a mutation in their genomic sequence, it can have severe consequences for the entire organism, such as abnormal development or cancer. In a healthy organism, cell signalling networks are only activated by signalling molecules if needed and secured by feedback mechanisms. For example, certain cells in the skin would receive a signal to grow and divide in order to close a wound. Once the wound is closed, the signal would cease and cells consequently stop dividing. In cancer cells however, cell signalling networks are often permanently activated and uncoupled from any effector molecules or feedback mechanisms (Hanahan and Weinberg 2011), consequently leading to activation of different gene expression programs and eventually causing indefinite cell divisions limited only by the extent of metabolic supplies or countermeasures of the body's immune system.

In melanoma, a rare but malign type of skin cancer, the *BRAF* gene commonly harbours a particular point mutation which leads to constitutive activation of its gene product, the BRAF protein (Brose et al. 2002). In its mutated form BRAF no longer depends on growth factor–induced signalling, no longer responds to feedback elements, and hence reshapes gene expression to ultimately initiate oncogenic transformation of melanocytes into invasive melanomas (Shain and Bastian 2016). From a historic point of view on cancer research and cancer etiology, it is worth mentioning the initial discovery of RAF proteins, the protein family of signalling elements which BRAF belongs to and which are all part of the ERK signalling network introduced in the previous section. Like many other oncogenes, RAF

proteins were first identified in retroviruses (Rapp and Goldsborough 1983). In the 1960s and 1970s, it was widely assumed that cancer was caused by viral infections (Blackadar 2016). The identification of retroviral oncogenes like *v-src*, *v-ras*, *v-erbB*, and *v-raf* (Duesberg and Vogt 1970; Shih et al. 1979; Bister and Duesberg 1979; Rapp and Goldsborough 1983) initially supported this conjecture as they were all capable of transforming host cells into cancer cells. But shortly after, the cellular origins of these oncogenes were discovered (Stehelin et al. 1976; Der, Krontiris, and Cooper 1982; Downward et al. 1984; Bonner, Kerby, and Sutrave 1985). Although ridiculed at first, it became more and more evident that the retroviral oncogenes were mere derivatives of their cellular counterparts alienated over the course of evolution to exploit the host cell machinery for viral reproduction (for a historical primer, refer to Vogt 2012).

Today, it is widely accepted that human cancers are most often caused by genomic alterations rather than by viral infections (Vogelstein et al. 2013), with a few prominent exceptions like human papillomavirus (HPV)–induced cervical cancer (Walboomers et al. 1999). Yet, many of the oncogenes initially discovered in retroviruses remain at the center stage in cancer research as their cellular counterparts were later found to be part of signal transduction networks commonly altered in cancer. Among these, the aforementioned ERK signalling network is of particular interest with many of its components being involved in cancer initiation and progression (Dhillon et al. 2007). Across publicly available cancer studies, 62.7% of patients carry alterations in at least one of the elements of the ERK signalling network.[4] Here, the top most frequently altered ones are signalling proteins KRAS, BRAF and EGFR (in approximately 14.1%, 8.9% and 7.1% of cases respectively). When altered, they acquire the potential to promote tumour cell growth or survival as elaborated for the BRAF case.

Certain cancer entities are particularly enriched for ERK signalling–related alterations. In case of BRAF, it is melanoma, thyroid cancer and pilocytic astrocytoma patients who commonly carry genomic alterations. Among all cancer entities, the highest BRAF alteration frequency was reported for pilocytic astrocytoma, a rare brain cancer with only 3.000 new cases diagnosed per year (US). Here, 85.4% of patients bear BRAF mutations. For melanoma, about 59.2% of patients carry BRAF mutations. This rate is similar to patients with thyroid cancer (58.7%), but survival rates differ dramatically. In case of thyroid cancer, 84.8% of patients survive five years after initial diagnosis, whereas only 53.2% of melanoma patients survive this period. Survival is even worse for melanoma patients carrying an additional alteration in EGFR or KRAS. Only 24.8% of patients in this group survive a five-year period.

[4] All alteration frequencies and survival rates mentioned in the following paragraphs are based on a meta analysis of cancer studies publicly accessible via cBioPortal. Results are shown in fig. 1.2 and the performed analysis is described in sec. A.1.

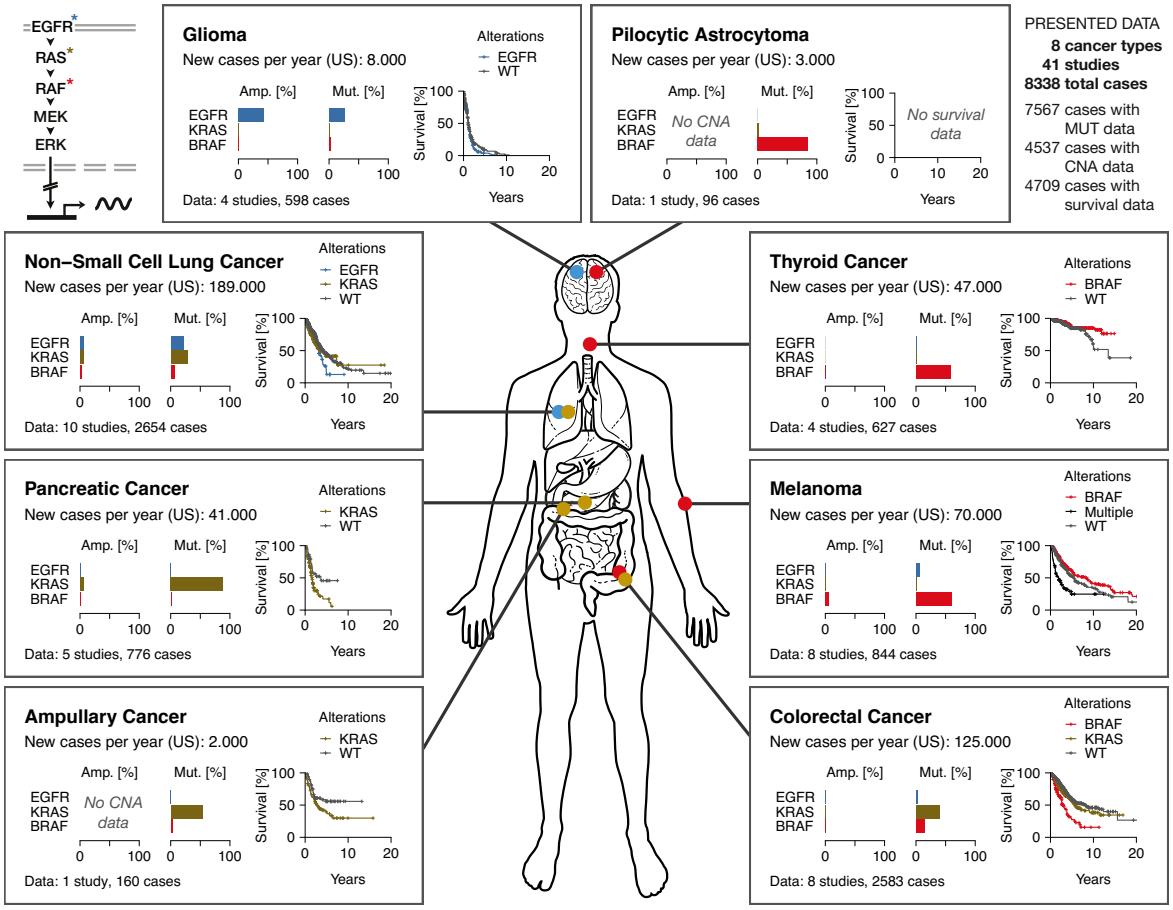


Figure 1.2 ERK signalling network alterations are present in many different cancer entities.
This figure summarises mutation and amplification frequencies of ERK signalling network elements EGFR, KRAS and BRAF as well as Kaplan-Meier curves across eight different cancer entities. Mutation status, CNA status and survival data was extracted from cBioPortal (Gao et al. 2013). New cases per year estimates were based on age-adjusted incidence rates for US population (CDC 2018). In Kaplan-Meier curves, WT (wild-type) cohorts include patients with no alterations in any of the three tested genes. For melanoma, *Multiple* refers to cohort of patients with at least two alterations in these genes. All other cohorts include patients with alterations in the indicated gene but without alterations in any of the two other genes. A simplified abstraction of the ERK signalling network is shown in the top left corner of the figure. Summary statistics for all considered cases to produce this figure are listed in the top right corner. For more information on this analysis, refer to sec. A.1 and supplementary tables B.1 and B.2. Human anatomy image source: Kirill Kazachek, published under flaticon basic license.

KRAS alterations are most commonly found in pancreatic cancer, ampullary carcinoma, colorectal cancer and non-small cell lung cancer (NSCLC). Indeed, 88.6% of pancreatic cancer patients and 53.8% of ampullary carcinoma patients carry KRAS mutations. In both cases, prognosis is worse for patients tested positive for KRAS alterations compared to patients

who still possess an intact (wild-type) form of KRAS. For NSCLC and colorectal cancer the situation is quite different. Here, KRAS mutation status does not significantly affect patient survival. Also, NSCLC and colorectal cancer patients may carry alterations in EGFR or BRAF respectively, which is rarely the case in pancreatic cancer and ampullary carcinoma.

Lastly, EGFR alterations are most common in glioblastoma multiforme, an aggressive brain tumour with poor patient survival rates (8.7% five-year survival). Here, both EGFR amplifications and EGFR mutations occur at high rates (43.8% and 25.4%) but do not significantly affect prognosis. Although neither EGFR mutations nor copy number alterations are prevalent in colorectal cancer patients, EGFR still plays an important role in colorectal cancer, as EGFR expression levels are often elevated (Porebska, Harlozińska, and Bojarowski 2000).

Whereas the relevance of BRAF, KRAS or EGFR alterations for tumour growth and survival became evident shortly after their initial discovery in retroviruses in the 1980s, it took decades until these insights could be translated to the clinic. Even today, classical approaches like chemotherapy, radiotherapy and surgery remain standard care in cancer therapy, but are now often complemented with so called targeted therapies (Druker 2003; Maughan 2017). Targeted therapies stratify patients not only by cancer type and stage, but consider patient's mutation or amplification status for oncogenes like EGFR, KRAS or BRAF. Drugs developed in this framework are tailored to specifically block oncogenic isoforms of these proteins. This is opposed to chemotherapy and radiotherapy, where treatment may likewise effect healthy cells potentially causing severe side effects. Also, targeted therapies commonly show higher efficacy than chemotherapy and have demonstrated the potential to entirely revert disease progression where classical approaches had only limited success. Here, the most prominent example is a drug called Imatinib which, at the same time, sparked the field of targeted therapies (Mauro and Druker 2001). Imatinib targets the oncogenic fusion protein BCR-ABL in patients suffering from chronic myelogenous leukemia (CML). Remarkably, the introduction of Imatinib as standard care for CML increased five-year survival from 30% to 86% (Druker et al. 2006) and hence encouraged development of targeted therapies for other cancer entities.

Despite this major improvement, the promises of targeted therapies were widely disillusioned by their overall moderate success (Maughan 2017). For example, initial efforts to develop drugs against EGFR, KRAS or BRAF in order to tackle aberrant ERK signalling in cancer were disappointing with no effect or only partial responses in patients (Downward 2003). Whereas patient responses to EGFR inhibitors Gefitinib and Erlotinib in NSCLC were later improved by mutation status–based stratification (Lynch et al. 2004; Tsao et al. 2005), no targeted inhibitor against the seemingly undruggable KRAS protein has entered the

clinic (Cox et al. 2014; Samatar and Poulikakos 2014). However, potential agents are finally in reach (Lito et al. 2016; Welsch et al. 2017). Regarding oncogenic BRAF, potent inhibitors were developed to either block all members of the RAF family (Sorafenib, cf. Escudier et al. 2007; Llovet et al. 2008) or to specifically block mutant BRAF (Vemurafenib and Dabrafenib, cf. Chapman et al. 2011; Hauschild et al. 2012). At first, patient responses to Vemurafenib and Dabrafenib were unprecedented, with near-complete regression in metastatic melanoma patients tested positive for mutant BRAF. Yet, patients commonly acquired resistance and suffered from severe disease relapse (Wagle et al. 2011; Bucheit and Davies 2013) once more dampening enthusiasm for targeted therapies.

Fortunately, novel concepts to overcome acquired resistance and to indirectly block activity of putatively undruggable proteins like KRAS are taking targeted therapy approaches to the next level (Samatar and Poulikakos 2014). Among the most promising ones are combination therapy and indirect blocking by targeting of upstream or downstream effectors. For instance, a combination of BRAF inhibitors Vemurafinib or Dabrafenib with an additional drug (Cobimetinib or Trametinib) targeting a different component of the ERK signalling network significantly improves overall survival of melanoma patients (Robert et al. 2015; Larkin et al. 2014). For colorectal cancer patients, the lack of potent KRAS inhibitors can potentially be circumvented by targeting EGFR and MEK in combination (Klinger et al. 2013; Misale, Arena, and Lamba 2014). Notably, application of EGFR-targeting drugs in colorectal cancer patients is independent of EGFR mutation status. Instead, stratification is based on KRAS mutation status and EGFR gene expression levels, emphasising that mere matching of oncogenic driver mutation status and available targeted therapies may not be sufficient in all cases. At the same time, it refutes the assumption that blocking of a signalling component should not be effective in a situation where tumour growth is driven by an oncogenic element downstream of that component.

Both combination therapy and indirect targeting follow a common theme. It is to not only consider patient mutation status, but to study and understand its effect on the entire signalling state of a cell. In this regard, previous research has demonstrated that changes in signalling network dynamics elicited by oncogenic drivers can be inferred from perturbation experiments and used as a rationale to suggest promising combination therapy candidates (Klinger et al. 2013). Since changes in signalling network dynamics propagate to changes in gene expression dynamics, one can hypothesise that studying gene expression dynamics may likewise help to identify more predictive biomarkers and potential drug targets in cancer. And indeed molecular classification of patient sub-groups based on genome-wide gene expression signatures was introduced shortly after the advent of high-throughput technologies (Ramaswamy et al. 2001; Veer et al. 2002). At the same time the idea of transcription

therapy was coined (Pandolfi 2001), an approach which entails the blockade of gene regulatory changes evoked by oncogenic signalling. Most commonly, this involves inhibition of transcription factors or epigenetic regulators (Pandolfi 2001; Yan and Higgins 2013; Bhagwat and Vakoc 2015). As two very common oncogenes, *MYC* and *TP53*, encode for transcription factors, transcription therapy approaches are not always designed to act in an indirect fashion by compensating aberrations in upstream signalling, but can likewise be designed to directly target oncogenic transcription factors.

Several agents following either a direct or indirect transcriptional therapy approach have entered clinical trials (Yan and Higgins 2013). Mechanistically these agents for instance interfere with DNA-binding capabilities of transcription factors by occupying binding sites in the DNA (Blume et al. 1991; Wu et al. 2005; Smolewski 2008) or by functioning as decoy molecules which mimic transcription factor binding sites (Mann 2005). For the latter, agents blocking activity of ERK signalling–dependent transcription factor AP-1 have been shown to inhibit EGFR-induced invasion in human squamous cell carcinoma cell lines (Shiratsuchi, Ishibashi, and Shirasuna 2002). Compared to agents targeting specific transcription factors, drugs blocking epigenetic regulators commonly affect gene regulation in a broader manner. Most prominently, histone deacetylase (HDAC) inhibitors extensively affect a cell's transcriptional landscape in a rather unspecific manner. Many novel components blocking different HDAC isoforms are currently tested in clinical trials spanning a large range of different cancer entities (West and Johnstone 2014). To be fair, transcription therapy approaches are one of many to potentially complement traditional chemotherapies or established targeted therapies. Others to be mentioned are immunotherapy (Vanneman and Dranoff 2012) and metabolic targeting (Pan and Mak 2007), but are not in scope of this thesis and thus are not discussed here.

In this section, it was shown that insights from cell signalling research enabled the advent of targeted therapies. Unfortunately, success stories like the one of Imatinib in CML form the exception rather than the rule. An effective and lasting control of cancer progression hence demands a thorough examination of cellular signalling dynamics and the gene regulatory changes which they evoke. New tailored combination therapies will hopefully result from these efforts. Potentially, they will not only aim for the blockade of other signalling elements but complement established chemotherapies and targeted therapies by tackling cancer cells from additional angles. As exemplified, this could include novel transcription therapies, an approach where the oncogene-mediated changes in gene expression are subject to treatment. For this, a better understanding is required of how the oncogenic cell fate information is encoded in signalling dynamics and decoded by gene expression dynamics.

1.3 Encoding and decoding of signalling dynamics

According to Sidney Brenner, one of the founding figures of molecular biology, biological systems should not be reduced to metabolic machines converting energy and transforming molecules. Instead, he stresses that molecular biology introduced the notion of information into biology with nucleic acid sequences being its material basis. Hence, biological systems should rather be perceived as systems that process and decode molecular information (Brenner 2010). Of course, the most prominent example of molecular information decoding is given by the central dogma of biology itself: Genetic information is encoded in the nucleic acid sequence of DNA, transcribed into RNA and translated into a chain of peptides that eventually shape a functional protein.

Whereas protein structure and function are determined by the coding DNA sequence, cellular fates are encoded in signalling networks and decoded by gene regulatory networks (fig. 1.3). More precisely, the identity, concentration or exposure time of an extracellular cue is encoded in features of signalling dynamics like signalling amplitude, duration or frequency (Behar and Hoffmann 2010). In turn, the information contained in signalling network dynamics is interpreted by gene regulatory networks culminating in a cellular phenotype. Remarkably, in this process quantitative features of signalling dynamics are converted to qualitative 'all-or-none' cell fate decisions, for instance, whether to divide or not (Kolch et al. 2015).

At this point, it is important to note that signalling pathways were once widely perceived as a mere "collection of linear information transporting pipelines" (Kholodenko, Hancock, and Kolch 2010) which consistently transmit extracellular signals to cell fate specific genes. For example, certain pathways would sense stress signals and activate genes promoting cell arrest, and others would sense growth signals and activate genes promoting cell division. Yet, it was observed that different cellular phenotypes could be caused by the same signalling pathway and, vice versa, the same cellular phenotype could be caused by activation of different pathways. Thus, signalling pathways are now viewed as interconnected signalling networks in which transmission of information is not limited to signalling component identity, but can be likewise encoded in the temporal or spatial dynamics of signalling (Kholodenko, Hancock, and Kolch 2010; Purvis and Lahav 2013).

Examples for information encoding in features of signalling dynamics like amplitude, duration or frequency are numerous and have been catalogued in a series of reviews (Murphy and Blenis 2006; Kholodenko 2006; Kholodenko, Hancock, and Kolch 2010; Behar and Hoffmann 2010; Purvis and Lahav 2013; Kolch et al. 2015). On the contrary, the process of signal decoding has not been studied extensively and molecular mechanisms of decoding remain to be established (Blüthgen and Legewie 2008; Purvis and Lahav 2013). In the following,

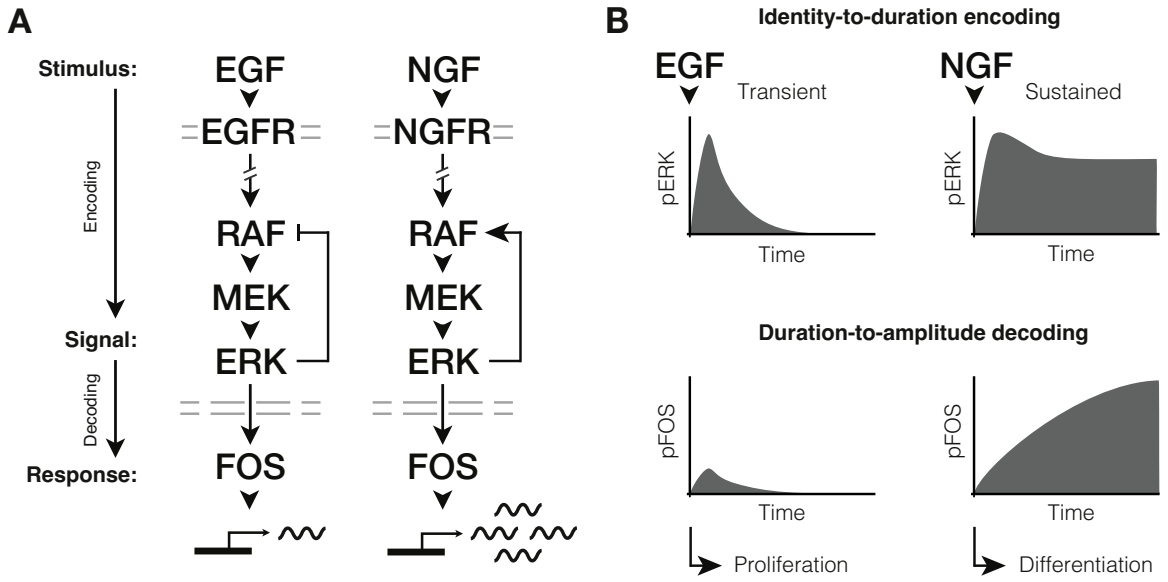


Figure 1.3 ERK signal duration decoding and the FOS protein sensor model.
A: An extracellular stimulus (EGF/NGF) is encoded in signalling network dynamics (ERK) and decoded by a molecular sensor (FOS). In PC12 cells, EGF-mediated signalling involves a negative feedback from ERK to RAF (left). In NGF-mediated signalling a positive feedback from ERK to RAF is caused by parallel activation of protein kinase C (right).
B: Schematic phospho-ERK and phospho-FOS levels in response to EGF and NGF in PC12 cells. Upper panel: The different network topologies shown in A allow for encoding of stimulus identity into signal duration. The negative feedback loop in EGF-mediated signalling causes transient ERK activity, whereas the positive feedback loop in NGF-mediated signalling causes sustained ERK activity. Lower panel: Transient ERK signalling is insufficient for stabilisation of FOS protein via phosphorylation and results in low levels of phospho-FOS (pFOS). Only sustained ERK signalling allows for accumulation of FOS protein due to phosphorylation-mediated stabilisation. According to the FOS protein sensor model (Murphy et al. 2002), accumulation of FOS protein eventually results in cellular differentiation of PC12 cells. The figure was freely adapted from Behar and Hoffmann (2010) and from Purvis and Lahav (2013).

I will discuss a selection of known signal encoding mechanisms and briefly touch upon the emerging efforts to study signal decoding mechanisms. To discriminate between the different features being encoded and decoded, it is helpful to follow a simple but comprehensive terminology proposed in one of the mentioned reviews (Behar and Hoffmann 2010). In short, one can classify encoding and decoding mechanisms into different types of stimulus-to-signal encodings and different types of signal-to-response decodings. For instance, stimulus-to-signal encoding mechanisms could be identity-to-duration or dose-to-frequency encodings and signal-to-response decoding mechanisms could be duration-to-amplitude or frequency-to-amplitude decodings.

One of the earliest described encoding mechanisms concerns ERK signalling in rat PC12 cells (Traverse et al. 1992; Marshall 1995). Here, treatment with epidermal growth factor (EGF) triggers cellular proliferation, whereas treatment with neuronal growth factor (NGF) triggers cellular differentiation (fig. 1.3). Based on this observation, one might speculate that the different growth factors activate different signalling pathways to trigger these phenotypes. However, it turns out that the ERK signalling network is activated in both scenarios but with different signal durations. Whereas EGF elicits transient ERK signalling, NGF elicits sustained ERK signalling. As the identity of these growth factors is converted into the duration of ERK signalling this process can be classified as identity-to-duration encoding. Mechanistically, the different signalling patterns are caused by different signalling network topologies (Santos, Verveer, and Bastiaens 2007). Upon EGF strong negative feedbacks from ERK to RAF and other elements of the network ensure transient ERK activity. On the contrary, NGF-mediated ERK signalling is sustained due to presence of a positive feedback from ERK to RAF via parallel activation of protein kinase C (PKC). Interestingly, signal durations and cellular fates could be reversed by reversing signalling network topologies (Santos, Verveer, and Bastiaens 2007) or by modulating growth-factor treatment frequency (Ryu et al. 2015). When promoting a positive feedback from ERK to RAF with help of a specific drug (phorbol 12-myristate 13-acetate, PMA), EGF treatment results in sustained ERK activation and cellular differentiation. Accordingly, inhibition of the positive feedback with a specific inhibitor (Gö7874) results in transient ERK activation and cellular proliferation upon NGF treatment (Santos, Verveer, and Bastiaens 2007). Likewise, repeatedly applied EGF treatment elicits high frequency pulses of ERK activity and can ultimately rewire cell fate to trigger differentiation in PC12 cells (Ryu et al. 2015).

Another example of identity-to-duration encoding was observed in 3T3 fibroblasts were growth factor identity is also converted into ERK signal duration (Murphy et al. 2002). Here, EGF elicits transient ERK signalling and no fibroblast proliferation whereas PDGF elicits sustained ERK signalling and S-phase entry. Notably, this study also proposes a mechanism for signal duration decoding, the FOS protein sensor model. According to the model ERK-mediated phosphorylation of FOS protein leads to its stabilisation and thereby causes protein accumulation. Accumulated FOS protein in turn is capable of inducing the expression of proliferative target genes. It was later suggested that the FOS protein sensor model of duration-to-amplitude decoding could also explain expression of prodifferentiation genes in PC12 cells upon NGF treatment (Pellegrino and Stork 2006).

In addition to the observations in PC12 cells and 3T3 fiborblasts, previous research has also reported identity-to-duration encoding in human breast cancer MCF7 cells. Here, transient signalling caused by EGF leads to proliferation, whereas sustained signalling caused by

heregulin (HRG) leads to differentiation (Nagashima et al. 2006). Similarly, hamster lung fibroblasts (CCL39) and mouse hippocampal cells (HT22) can discriminate transient and sustained ERK signalling. In both cell types only prolonged ERK activity accompanied by ERK nuclear retention causes cell death, whereas transient nuclear translocation of ERK is insufficient to do so (Lenormand et al. 1998; Stanciu and DeFranco 2002).

Interestingly, a series of additional examples comes from different studies on cells of the immune system. For example, in T cell development, ERK signalling is required for differentiation of progenitor thymocytes into CD4 T cells. It was initially reported that the strength of the signal corresponds to the number of cells differentiating into CD4 T cells (Sharp et al. 1997). However, it was later found that ERK signal duration is likewise important for T cell differentiation, since positive selection of T cells in the thymus requires sustained ERK signalling (Mariathasan et al. 2001). Transient signalling on the contrary leads to negative selection. In a different study, it was found that human chronic myelogenous leukemia cells (K562) undergo megakaryocytic differentiation only in response to sustained ERK signalling elicited by exposure to PMA (Whalen et al. 1997). Lastly, a special case of identity-to-duration encoding was described for NF-κB signalling, where different inflammatory agents can cause different activity patterns in various mammalian cell lines. Upon TNFα treatment multiple short pulses of NF-κB activity are observed, whereas LPS treatment causes sustained NF-κB activity (Hoffmann et al. 2002). Closely related to this observation, stimulus-to-duration encoding was reported for Toll-like receptor 4 (TLR4) signalling in macrophages (Litvak et al. 2009). In these cells, transient bacterial infections elicit transient signalling and no inflammatory response, whereas persistent bacterial exposure elicits sustained signalling and the transcription of genes critical for protective inflammatory responses. To be correct, one should not consider the stimulus to be encoded in this case, as the sensed input feature (stimulus duration) is only relayed to the signalling network (signal duration).

Although all studies discussed so far focused on different mammalian model systems, signal duration encoding and decoding has likewise been described in less complex organisms. In Xenopus oocytes, only high doses of progesterone activate MAP Kinase signalling and are required for subsequent maturation (Huang and Ferrell 1996; Ferrell and Machleder 1998). For yeast pheromone signalling it was shown that pheromone dose is encoded in signal duration (Behar et al. 2008). This dose-to-duration encoding allows downstream duration-to-amplitude decoding and either results in growth or mating of yeast depending on whether or not a slowly accumulating decoder surpasses a certain threshold (Behar et al. 2008). On top of the variety of stimulus-to-duration encoding examples from mammalian and non-mammalian systems enumerated here, there is a multitude of other encoding mechanisms that have been reported in the literature. These include dose-to-frequency and dose-to-

amplitude encoding in yeast glucose sensing, as well as a complex stimulus dose and identity to signalling amplitude and frequency encoding in p53 signalling (Hao and O'Shea 2012; Batchelor et al. 2011; Loewer et al. 2010).

In contrast to the extensive literature on signal encoding in general, only little is known about signal decoding. Apart from the aforementioned FOS protein sensor model (Murphy et al. 2002), a few other feedforward motifs have been suggested to be capable of signal decoding. These include duration-to-amplitude decoding in PDGF-induced ERK signalling and aforementioned LPS-induced TLR4 signalling, as well as frequency-to-amplitude decoding in p53 signalling (Toettcher, Weiner, and Lim 2013; Litvak et al. 2009; Porter, Fisher, and Batchelor 2016). In case of PDGF-induced ERK signalling, a second candidate was identified to potentially decode sustained ERK activity in addition to FOS protein (Toettcher, Weiner, and Lim 2013). It was observed that only prolonged signalling allows for ERK nuclear translocation and subsequent STAT3 phosphorylation. Phosphorylated STAT3 is then secreted and functions in a paracrine circuit to activate IL-6 family receptors of neighbouring cells. In TLR4 signalling transcription factors NF-κB and C/EBPδ form a coherent feedforward motif (Litvak et al. 2009). As a result, C/EBPδ only accumulates in response to sustained bacterial exposure and sustained NF-κB activity, ultimately leading to an innate immune response in macrophages. Frequency-to-amplitude decoding in p53 signalling was observed in a recent single-cell analysis of MCF7 breast cancer cells (Porter, Fisher, and Batchelor 2016). Notably, the identity of the decoder/sensor in this case is not a protein, but a set of long-lived mRNAs. Here, a low frequency of p53 signalling pulses in response to DNA double strand breaks is insufficient for accumulation of these mRNAs. Only high frequency pulses of p53 signalling upon periodic Nutlin-3 treatment allow for accumulation of these targets and result in a more densely connected gene regulatory network.

In this section it was shown that different properties of extracellular cues, such as stimulus dose or identity, can be encoded in different features of signalling dynamics, such as signal duration or frequency, and decoded by particular response elements to eventually culminate in different cellular phenotypes. As many different signalling networks encode the dose or identity of extracellular cues in signal duration, the subsequent cellular task of signal duration decoding is of particular interest. In fact, signal duration decoding not only matters for cell fate decisions in healthy cells, as oncogenic signalling in cancer cells can often be considered as an extreme case of sustained signalling, again emphasising the need for a more detailed understanding of signal duration decoding in general and of ERK signal duration decoding in particular.

1.4 Scope of this study

Decades of research have provided a thorough understanding of the ERK signalling network and shed light on its ambiguous role in normal development and aberrant transformation of cells. On the one hand, it was found that its complex signalling dynamics are capable of controlling a multitude of phenotypic responses essential to normal development of organisms of all kinds. On the other hand, its role in cancer initiation and progression has urged researchers to translate gained insights into viable treatment options. Both ways, the significance of ERK signalling for health and disease has motivated detailed models to mechanistically understand how the information flow from extracellular cues to cell fate decisions materialises at the molecular level. In this regard, the translation of stimulus and signalling input dynamics to target gene output dynamics has been a key subject in many different studies. As outlined above, previous efforts often focused on the examination of signal encoding, whereas the identification of mechanisms that decode signalling dynamics was coined to remain "one of the most challenging goals for the field" (Purvis and Lahav 2013). In my thesis, I now provide an extensive analysis of the relation between ERK signalling dynamics and subsequent changes in gene expression dynamics to gain new insights on the fundamental process of ERK signal duration decoding.

In the first part (ch. 2), different potential experimental models are discussed to study mRNA dynamics and ERK signalling, and a mathematical model is introduced to complement empirical data with theoretical considerations. After an initial discourse of different model systems, one particular system that allows for tight control of ERK signal duration is introduced (HEK293ΔRAF1:ER). Both genome-wide gene expression data and ERK activity data acquired from this system are presented. Based on a translational shut-down approach, a comprehensive catalogue of early (primary) and late (secondary) response genes is defined. Next, the aforementioned mathematical model of gene expression is formulated to simulate potential gene regulatory strategies in response to ERK signalling. Here, it is examined how different model parameters shape mRNA dynamics of PRGs and predictions are made to identify which dynamic features potentially govern ERK signal duration decoding.

In the second part (ch. 3), the introduced mathematical framework is trained on a subset of the acquired gene expression and ERK activity time-course data to gain gene-wise parameter estimates for transcription and degradation rates. Based on these parameter estimates, all responding PRGs are classified to distinguish IEGs and DEGs as well as to identify a new group of PRGs termed immmediate-late genes (ILGs). The model is further validated by semi-quantitative prediction of gene expression dynamics in response to both synthetic and physiological ERK signalling input dynamics. Next, a genome-wide quantification of mRNA half-lives is performed with help of two independent methods that allow for direct

and indirect measurements of mRNA half-life. For direct measurements time-course data is obtained from transcriptional shut-down experiments by using the transcriptional inhibitor Actinomycin D (ActD) to directly trace mRNA decay over time. A thiol-labelled uridine analogue (4SU) is used for RNA metabolic pulse labelling experiments to indirectly infer mRNA half-lives from the ratio of pre-existing (unlabelled) and newly synthesised (labelled) mRNA. All half-life estimates are later validated by comparison to previously published data sets on mammalian mRNA half-lives and an evaluation of the different methods is provided. Lastly, in a comparative analysis, mRNA half-lives, transcription rates and gene promoter GC content are studied across the identified gene expression modules for a more detailed characterisation of these modules.

The third and final part (ch. 4) focuses on examining the capability of the identified gene expression modules to decode ERK signal duration. The relation between signal duration and response amplitude is investigated by comparing the behaviour of IEGs and ILGs in response to a wide range of different signal durations. First, both mRNA and protein levels are quantified for a subset of genes using real-time quantitative PCR (RT-qPCR) and western blotting. Next, a genome-wide analysis of approximately six hundred responding target genes is performed to systematically investigate the role of mRNA half-life in the process of duration-to-amplitude decoding. The relation between ERK signal duration and response amplitude is further analysed in PC12 cells and MCF7 cells, two paradigm model systems for ERK signal duration decoding. Lastly, the molecular function of IEGs, DEGs and ILGs is assessed with help of Gene Ontology (GO) enrichment analysis and the results are related to the observed cellular phenotypes of HEK293ΔRAF1:ER cells in response to transient versus sustained ERK signalling.

The three main chapters are followed by a comprehensive conclusion where I summarise and discuss all findings presented in my thesis. The subsequent appendix lastly includes a material and methods section which provides supplementary information on all analyses presented in this thesis including both experimental procedures and computational methods, as well as supplementary tables and indices of bibliographical references, acronyms and abbreviations and a list of figures.

2 Experimental and mathematical models to study mRNA dynamics upon ERK signalling

Essentially, all models are wrong, but some are useful.
—George E.P. Box (1987, p. 424)

2.1 Introduction

Biology started off as a descriptive science with its main goal to catalogue nature and classify species. With the introduction of mathematical models into the different branches of biology –like the renowned modelling of enzyme kinetics in biochemistry (Menten and Michaelis 1913), the modelling of population dynamics in ecology (Snider 2014; Volterra 1926), or the quantification of membrane potentials in neurophysiology (Goldman 1943; Hodgkin and Huxley 1952)– biology slowly began to turn from a qualitative discipline into a quantitative one. This transformation was sped up with the rise of molecular biology, when it became possible to quantify a range of cellular and subcellular processes, like protein biosynthesis and the underlying regulatory dynamics of gene expression (Monod, Pappenheimer Jr., and Cohen-Bazire 1952; Gorini and Maas 1957). The very notion of gene regulation was coined in 1961 when François Jacob and Jacques Monod, for the first time, described the existence of a gene regulatory element: the *lac operon* (Jacob and Monod 1961). At that time it was known that DNA, not protein, is the hereditary material (Avery, Macleod, and McCarty 1944; Hershey and Chase 1952), the genetic code was described later that year (Crick et al. 1961) and the adaptor molecule that translates genomic information into amino acid sequences had been discovered (Hoagland et al. 1958).

However, the central dogma of biology was yet incomplete, as it remained unknown how the structural information concealed in the nucleus could be transmitted to the cytoplasm, where protein biosynthesis was known to take place. In their landmark experiment, Jacob and Monod deciphered the nature of the hypothetical intermediate required for this transmission by studying lactose metabolism in *Escherichia coli*. They found that the synthesis

of proteins is mediated by short-lived messengers (mRNAs) that transcribe structural genes (genomic sequences) and transmit the information to the cytoplasm where they are translated with help of the previously identified adaptor molecules (tRNAs) into functional gene products (proteins). One must emphasise that Jacob and Monod immediately noted and experimentally validated that the messenger RNA is "a very short-lived intermediate both rapidly formed and rapidly destroyed during the process of information transfer" (Jacob and Monod 1961). Moreover, they identified its short lifespan as a crucial property required by the kinetics of gene induction.

Notably, Monod and colleagues studied the kinetics of gene induction even before identification of the messenger RNA. In 1952, they proposed a mathematical model of enzyme biosynthesis to simulate the dynamics of gene induction (Monod, Pappenheimer Jr., and Cohen-Bazire 1952). The model includes only two parameters, a production rate at which the gene product would be synthesised, and a degradation rate at which the gene product would decay (termed k and γ here):

$$\xrightarrow{k} \text{gene product} \xrightarrow{\gamma}$$

Initially, the model was developed to study the kinetics of protein biosynthesis and served as a reasonable abstraction to mathematically describe gene regulatory dynamics with an ordinary differential equation (ODE). However, this initial model did not yet account for the dynamics of the messenger RNA. Hence, it was later extended to consider both mRNA and protein dynamics in a system of coupled first-order ODEs or simply adopted to exclusively model mRNA dynamics without studying propagation to protein dynamics (reviewed in Yagil 1975 and Hargrove, Hulsey, and Beale 1991). Whereas early studies of gene expression modelling were limited to specific genes or to bulk estimates (Price et al. 1962; Singer and Penman 1973), high-throughput technologies like RNA microarrays and next generation sequencing (NGS) now allow to thoroughly investigate the global landscape of gene expression dynamics in the cell at steady state (Schwanhäusser et al. 2011) or upon different perturbations (Zeisel et al. 2011; Rabani et al. 2011; Rabani et al. 2014; Pretis et al. 2015). Studies in this field have identified different gene regulatory strategies. It was shown that for most genes up- or downregulation is governed by changes in mRNA transcription rates (Rabani et al. 2014; Pretis et al. 2015). Only a small fraction of genes is regulated by additional changes in processing or degradation rates.

Whereas these studies thoroughly examined gene regulatory strategies in a genome-wide manner, less effort was devoted to relate these dynamics to the signalling network dynamics they are evoked by. Instead, attribution was limited to stimulus identity. For example, gene regulatory dynamics have been investigated in response to a number of growth factors

including epidermal growth factor (EGF), fibroblast growth factor (FGF), neuronal growth factor (NGF), heregulin (HRG) and platelet-derived growth factor (PDGF), all of which are capable of activating ERK signalling. However, from such data it is difficult to address how different ERK signalling input dynamics are translated into gene expression output dynamics, as additional signalling networks can be activated by these growth factors (Kholodenko, Hancock, and Kolch 2010). The activated pathways can intertwine, counteract each other or converge on the same downstream promoters (Parikh et al. 2010). Negative feedback loops allow for adaptation to constant signal exposure and complicate the attempt to link gene expression programs to signalling inputs even further. As elaborated in the general introduction, growth factor–induced RAF-MEK-ERK and PI3K-AKT signalling networks cross talk extensively (Mendoza, Er, and Blenis 2011; Aksamitiene, Kiyatkin, and Kholodenko 2012; Fritsche-Guenther et al. 2016). They share downstream transcription factors (Tullai et al. 2004) and a multitude of feedback loops act on both pathways (Kolch et al. 2015). To overcome the apparent issue of signalling network attribution inherent to growth factor-induced systems, more complex experimental systems that allow for gene-specific perturbations need to be considered.

In this chapter, different model systems are discussed that are capable of specifically attributing gene regulatory output dynamics to cell signalling input dynamics. Subsequently, gene expression time-course measurements obtained from one of these systems will be presented. Eventually, a mathematical framework for modelling of gene expression dynamics will be introduced. Initially, the framework will be employed to simulate mRNA response dynamics upon different ERK signalling input scenarios, and, in the following chapter, it will be used for modelling and prediction of the acquired time course data and guide identification of different gene expression modules with distinct dynamic properties.

2.2 Experimental model systems to investigate gene-specific cellular responses

A variety of experimental model systems has been developed to study gene-specific cellular responses. For example, they can be used to distinctively characterise the gene expression program linked to a perturbed signalling protein. Depending on the controllability and induction kinetics of the system, they may even be suitable to study the regulatory effects of different signalling features like signal duration, amplitude or frequency. In general, experimental model systems that allow for gene-specific perturbations can be categorised along two axes (fig. 2.1A). First, one can discriminate systems where a certain gene product is activated and systems where a certain gene product is inactivated. Secondly, time scales can

be very different as perturbations may have an instant, slow or permanent effect.

Paradigmatic examples for permanently perturbed systems are knock-out cell lines or animal models where genome editing methods such as Cre-loxP (Sternberg and Hamilton 1981; Gu, Zou, and Rajewsky 1993) or CRISPR-Cas9 (Jinek et al. 2012) are employed to entirely disrupt the expression of a gene of interest by deleting parts of its promoter region or gene body from the genome. In recent years, many different variations of the CRISPR-Cas9 approach have been introduced beyond its application to generate knock-out model systems (Hsu, Lander, and Zhang 2014). For example, they allow to enhance or repress specific promoter sequences or to precisely exchange single nucleobases in the genome. Such methods also cause permanent perturbations, but the degree of activation or inactivation varies depending on the exerted method. When editing single nucleobases one might introduce gain of function (GOF) mutations, in which case a mutated gene is constitutively activated, or loss of function (LOF) mutations, in which case a mutated gene is permanently rendered inactive. For many cancer model systems, endogenous GOF and LOF alterations do not result from genome editing efforts, but are inherent to the model system and occured randomly in the patient or mouse model they were derived from.

When permanently perturbed systems are studied, it is often the case that differences between two steady states are assessed, i.e. the state of an unperturbed system and the state of its permanently perturbed derivative. In contrast, inducible systems allow to study the dynamic transition between an unperturbed state and a perturbed one as long as perturbation kinetics are sufficiently fast. For some inducible systems this is not the case. For example, drug-inducible transgene systems elegantly allow for overexpression of a gene of interest (Saez et al. 1997), but induction kinetics are most likely slower than the signalling dynamics of the perturbed network. Thus it is less common to study the temporal dynamics of the primary perturbation in such systems, but rather to test how an uninduced system (no transgene expression) behaves in comparison to a fully induced one (transgene overexpression) when both are exposed to a secondary agent. The same premise holds true for knock-down systems, where genes can be specifically downregulated in expression with help of RNA interference (RNAi) approaches (Hannon 2002). Similar to drug-inducible transgene systems targeting with shRNAs or siRNAs affects the expression levels of a gene of interest within a matter of hours. Whereas drug-inducible systems increase the expression of synthetically introduced transgenes, RNAi approaches reduce the expression of endogenous target genes. In both cases, the modification of gene expression levels serves as a proxy to modify gene product activity.

Whenever tight control of signalling dynamics at a time scale of seconds or minutes rather than hours or days is desired, it is necessary to directly modify the gene product activity

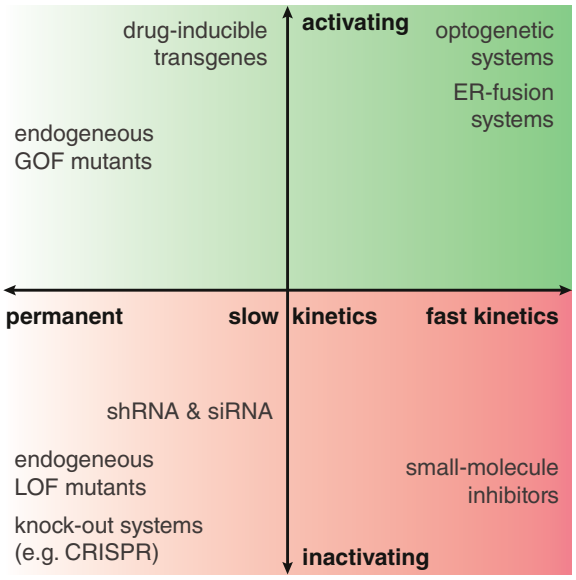


Figure 2.1 A selection of experimental model systems to study gene-specific cellular responses.
Many different experimental model systems can be used to study gene-specific effects on gene expression or cellular phenotypes. They can be divided along two axes. Here, the y-axis discriminates between activation and inactivation of a gene of interest. The x-axis shows temporal kinetics of the exerted perturbation. These can be fast in a range of seconds or minutes, rather slow in a range of hours or days, or even permanent. Please refer to the main text for a more detailed description of the mentioned systems and their benefits and limitations.

instead. This can be achieved in systems where a gene of interest is constitutively expressed, but remains in an inactive state unless an external stimulus is used to activate it. Two common types of systems that follow this design principle are optogenetic systems (Tischer and Weiner 2014) and oestrogen receptor (ER-)fusion systems (McMahon 2001). Both approaches have in common that the gene of interest is fused to an inducible protein that can either be controlled with light (optogenetic systems) or with oestrogen receptor ligands (ER-fusion systems).

In optogenetic systems the subcellular localisation of photosensitive proteins like Phytochrome B (PHYB) or Chryptochrome 2 (CRY2) is controlled with light pulses of defined excitation wavelengths. When a light pulse is presented, a conformational change of the photosensitive protein promotes the translocation of the fused gene of interest to or away from its normal site of action. In case of ERK signalling, different optogenetic systems have been established to control the activity of signalling proteins (reviewed in Tischer and Weiner 2014 and Zhang and Cui 2015). These include an optogenetic SOS:PHYB fusion system to ac-

tivate RAS via SOS:PHYB (optoSOS) membrane recruitment (Toettcher, Weiner, and Lim 2013; Wilson et al. 2017; Bugaj et al. 2018) and an optogenetic RAF:CRY2 fusion system to activate MEK via RAF:CRY2 membrane recruitment (Zhang et al. 2014; Wend et al. 2014).

The activation of hormone-dependent ER-fusion systems can be achieved with ER agonists like oestrogen or ER antagonists like 4-hydroxy-tamoxifen (4OHT). Here, 4OHT is usually favoured over agonistic ligands to prevent activation of endogenous ER signalling. Interestingly, the mechanism of activation is not fully understood for ER-fusion proteins. It is speculated, however, that fusion protein activity is sterically constrained by heat shock protein 90 (hsp90) binding in absence of ER ligands. When exposed to ER ligands, the hsp90 chaperone is removed from the fusion protein and in turn the fused signalling protein can bind to its substrates and activate them. Many different conditionally active ER-fusion proteins have been generated to study ERK signalling and have been introduced to a variety of cell lines. These include fusion proteins for RAS (ER:HRAS) (Tarutani et al. 2003), different RAF isoforms (ΔARAF:ER, ΔBRAF:ER, ΔRAF1:ER) and MEK (ΔMEK1:ER) (Samuels et al. 1993; McMahon 2001).

When comparing optogenetic systems and hormone-dependent ER-fusion systems, one of the key differences lies in their reversibility. For optogenetic systems, inactivation can be achieved by turning off the light source or by exposure to an inactivating wavelength. As a consequence the photosensitive proteins dissociate from the membrane within seconds (PHYB) or minutes (CRY2) (Tischer and Weiner 2014). This allows for complex signalling patterns like pulsatile signalling inputs (Wilson et al. 2017; Johnson, Shvartsman, and Toettcher 2018). To inactivate ER-fusion systems, the activating ligand has to be washed out from the system, or the system has to be inactivated with help of a small-molecule inhibitor that acts on one of the downstream signalling components.

Altogether, systems using conditionally active fusion-proteins should be favoured to study network–specific features of signalling dynamics like signal duration, amplitude or frequency. Hence, in my thesis, a well studied fusion protein (ΔRAF1:ER) is used to investigate ERK signal duration and its gene regulatory effects in human embryonic kidney (HEK293) cells. The following section will describe the model system and the different experiments performed to obtain transcriptome-wide gene expression time-course data in response to ERK signalling.

2.3 A highly controllable cell culture system to study ERK signalling dynamics

Most of the results presented in my thesis are based on experimental data obtained from a synthetic cell culture system that allows for tight control of ERK activity (Cagnol, Obberghen-

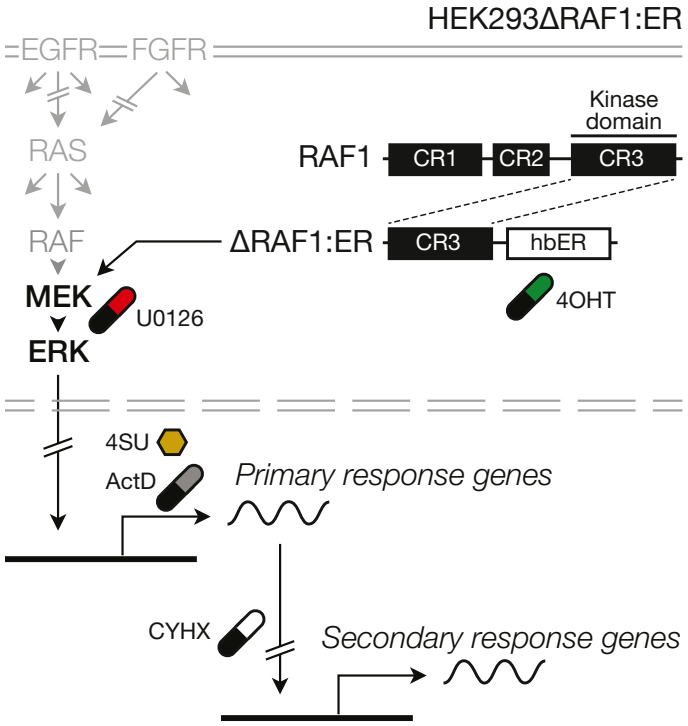


Figure 2.2 A synthetic model system to study ERK signal duration. HEK293 with a stably transfected ΔRAF1:ER fusion protein were treated with ER antagonist 4-hydroxytamoxifen (4OHT). To generate pulses, ERK signalling was turned off using the MEK inhibitor U0126. To identify primary response genes, translation was blocked with cycloheximide (CYHX) in parallel to 4OHT stimulation. In addition, Actinomycin D (ActD) was used to determine mRNA half-lives via transcriptional shut-down and 4-thiouridine (4SU) was used to determine mRNA half-lives via metabolic labelling.

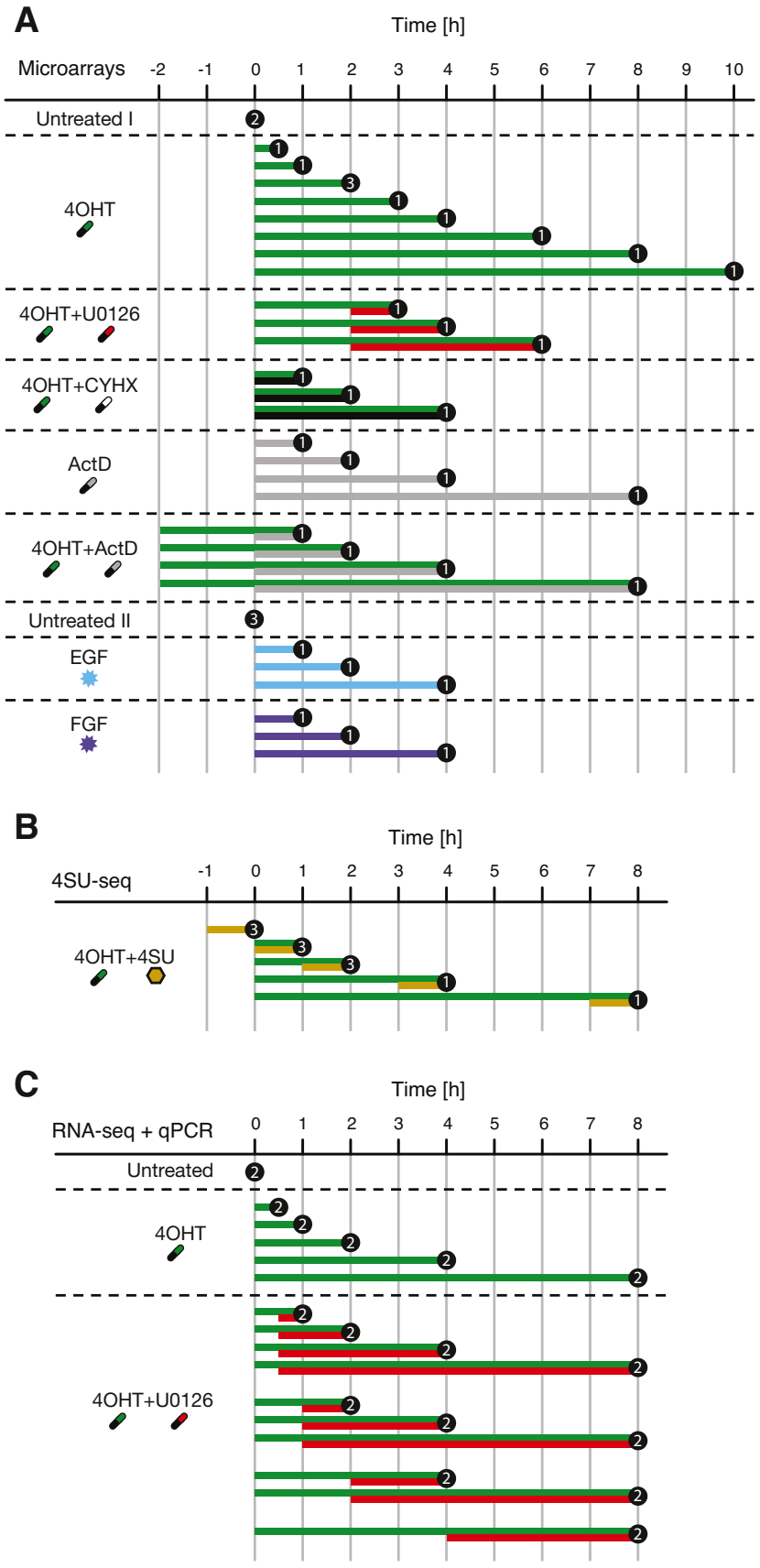
Schilling, and Chambard 2006). More precisely, HEK293 cells constitutively expressing an inducible form of RAF (ΔRAF1:ER) (Samuels et al. 1993; McMahon 2001; Cagnol, Obberghen-Schilling, and Chambard 2006) (fig. 2.2) were used to generate a wide range of different gene expression time course data sets (fig. 2.3). Here, ΔRAF1 refers to a truncated version of endogenous RAF1. It lacks RAF1 conserved regions CR1 and CR2 and only consists of its kinase domain (CR3). CR1 and CR2 contain different functional domains (Lavoie and Therrien 2015) and their absence in ΔRAF1 has several implications for ΔRAF1 regulation. CR1 contains a RAS-binding domain (RBD) and a cystein-rich domain (CRD). It has an auto-inhibitory effect on the kinase domain which is only disrupted via RAS binding. CR2 contains and is surrounded by several Ser/Thr sites, many of which have been described as negative feedback sites regulated by ERK (Fritsche-Guenther et al. 2011) and other proteins (Dhillon et al. 2002). Hence, ΔRAF1 is uncoupled from RAS signalling, it does not have

any auto-inhibitory activity and it is not affected by negative feedback regulation. Instead, its activity can be experimentally controlled when fused to the hormone binding domain of oestrogen receptor (hbER) as described for ER-fusion systems in the previous section. In contrast to growth factor-induced systems, ERK signalling via ΔRAF1:ER avoids pathway divergence and distinctively allows attribution of measured downstream effects to the RAF-MEK-ERK signalling cascade.

When ΔRAF1:ER is introduced into HEK293 cells, a constant exposure to 4OHT causes sustained activation of ERK signalling and results in caspase-8-mediated induction of apoptosis, whereas parallel treatment with small molecule MEK inhibitor U0126 does not (Cagnol, Obberghen-Schilling, and Chambard 2006). In addition, sustained ERK signalling in this system mimics oncogenic RAF signalling and therefore can provide insights into early onset of RAF-driven malignancies and the decisive competition between anti-apoptotic and pro-apoptotic signals elicited by the RAF-MEK-ERK signalling network. For my thesis, the HEK293ΔRAF1:ER model system was used to generate a range of ERK signalling pulses with defined duration, by stimulating ΔRAF1 activity with 4OHT and subsequently blocking it using U0126. In total, eight different treatments or combinations of treatments were performed including transcriptional and translational shut-down experiments. Three independent experimental methods were used to measure mRNA expression levels, namely microarrays, RNA-sequencing and quantitative real-time PCR (RT-qPCR).

First, multiple series of microarray time-course expression data sets were obtained (fig. 2.3A). Sustained ERK signalling was profiled for eight time points ranging between thirty minutes and ten hours upon 4OHT treatment. A two-hour pulse of ERK signalling was profiled at three additional time points after subsequent treatment with U0126. Parallel treatment with CYHX and 4OHT for a period of one, two and four hours was performed to identify a sub-

Figure 2.3 ***(facing page)*** **Acquired samples from HEK293ΔRAF1:ER cells.**
A: Microarray time course data was used for model fitting and cluster definition. Numbers indicate independent biological measurements per condition and time point. All microarray experiments were performed by Raphaela Fritsche-Günther, Anja Sieber and Nadine Lehmann.
B: 4SU-sequencing time course data of metabolically labelled (4SU) cells was used for determination of transcription rates and steady-state half-lives in untreated cells. All 4SU-sequencing experiments were performed by Anja Sieber and Emanuel Wyler.
C: RNA-sequencing and RT-qPCR time course data was used for validation of the identified signal duration decoding principle. All RNA-sequencing and RT-qPCR experiments were performed by myself.
For all panels, each row corresponds to one sample and numbers indicate independent biological replicates acquired per sample. Treatment concentrations are provided in section A.2.


A
Time [h]
Microarrays
-2 -1 0 1 2 3 4 5 6 7 8 9 10
Untreated I
4OHT
4OHT+U0126
4OHT+CYHX
ActD
4OHT+ActD
Untreated II
EGF
FGF
B
Time [h]
4SU-seq
-1 0 1 2 3 4 5 6 7 8
4OHT+4SU
C
Time [h]
RNA-seq + qPCR
0 1 2 3 4 5 6 7 8
Untreated
4OHT
4OHT+U0126

set of primary response genes. Two different time series were acquired that involved ActD treatment. ActD-mediated transcriptional shut-down was performed to allow assessment of mRNA degradation in uninduced cells, whereas prior treatment with 4OHT was performed for a period of two hours to assess potentially altered mRNA degradation in response to ERK signalling. In addition to these drug-induced perturbations, two microarray time series of physiological treatments with growth factors EGF and FGF were acquired. All drug-induced perturbations were obtained in a first batch of microarray hybridizations including three independent untreated control samples, one of which was later excluded from the analysis due to a suspected bacterial contamination (cf. sec. A.3). All growth factor–treated samples were obtained in a second batch including three independent untreated control samples. Secondly, a time course of samples labelled with 4SU was acquired to enable the assessment of mRNA transcription rates in response to ERK signalling and to validate mRNA half-life estimates derived from ActD experiments (fig. 2.3B). For this, up to three different mRNA fractions were obtained containing either nascent (labelled), pre-existing (unlabelled) or total (labelled plus unlabelled) RNA and subsequently subjected to RNA-sequencing (cf. sec. A.4). Lastly, an additional genome-wide set of gene expression time-course data was generated to further investigate the role of mRNA half life in ERK signal duration decoding (fig. 2.3C). Here, five different pulses of ERK signal duration were profiled with help of 4OHT treatment and subsequent U0126 treatment. Initially, these samples were only subjected to RT-qPCR for a selected number of genes, but were later also profiled with help of RNA-sequencing (cf. sec. A.4 and A.8). Altogether, a range of different transcriptomics time-course data sets was obtained from HEK293ΔRAF1:ER cells to thoroughly investigate the relation between ERK signal duration and downstream mRNA dynamics and to study the role of mRNA half-lives in this regard. The next section examines a subset of the initially obtained microarray data to profile the gene regulatory response to ERK signalling. All other data sets are discussed in chapters 3 and 4 of my thesis.

2.4 Profiling the gene regulatory response to ERK signalling

Previous studies have reported that ERK signalling induces a multitude of early and late responding genes (Amit et al. 2007; Tullai et al. 2007; Dijkmans et al. 2009; Saeki et al. 2009; Nagashima et al. 2009; Stelniec-Klotz et al. 2012). To summarise, primary response genes (PRGs) are induced which in turn mediate expression of secondary response genes (SRGs) (Yamamoto and Alberts 1976). This dependency delays SRG induction and at the same time allows for decoding of signal duration, as only prolonged ERK activity ensures sufficient production of required primary factors. In accordance, primary response gene and transcription factor FOS was described to function as a molecular sensor for ERK

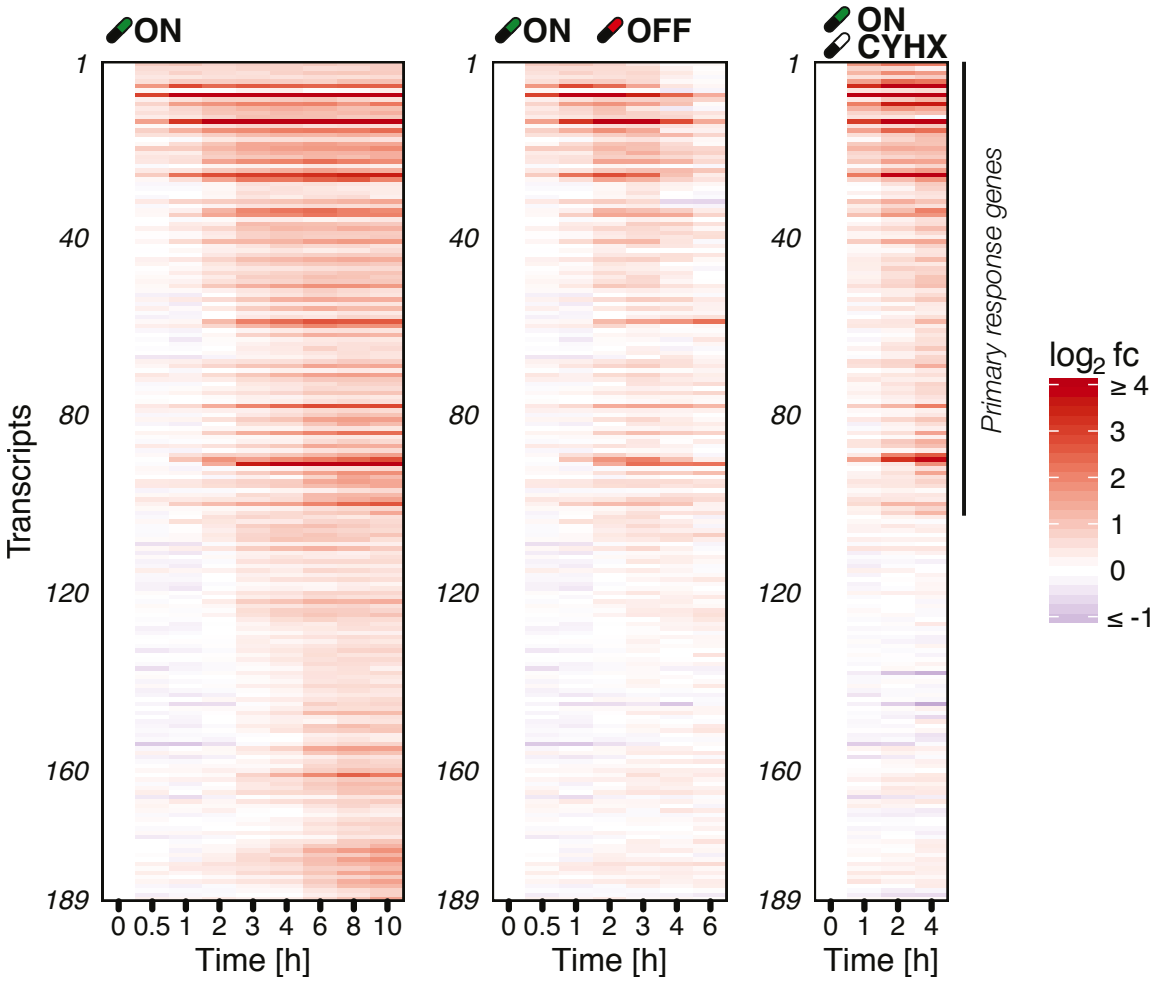


Figure 2.4 Expression kinetics in HEK293ΔRAF1:ER cells.
Log_2 gene expression fold changes of 189 significantly induced genes (FDR = 1%) across different treatment scenarios. Gene induction of immediate, delayed and late responding genes is sustained upon constant activation (ON scenario) and transient upon two-hour pulse activation (ON/OFF scenario). Genes significantly induced upon parallel CYHX treatment were considered primary response genes (indicated by black line). Genes were ranked by their model-derived response time (cf. sec. 3.2). Green icon indicates 4OHT treatment, red icon indicates subsequent U0126 treatment and white icon indicates parallel CYHX treatment.

signal duration (Murphy et al. 2002; Murphy, MacKeigan, and Blenis 2004). When ERK signalling is sustained, FOS protein is stabilised and can promote transcription of specific SRGs. In contrast, when ERK activity is transient, its signal declines before FOS protein can accumulate (Whitmarsh 2007).

However, the protein sensor model and the concept of PRGs and SRGs cannot explain the ample observation of late primary response genes. When induction of SRGs is blocked with

help of the protein biosynthesis inhibitor CYHX, not only immediate-early primary response genes (IEGs) have been found differentially regulated, but also an extensive set of delayed primary response genes (DEGs) (Amit et al. 2007; Tullai et al. 2007). IEGs peak about 30 minutes post EGF stimulation and are succeeded by DEGs, peaking about 120 minutes post stimulation (Avraham and Yarden 2011). Both IEGs and DEGs are primary response genes (PRGs), as they do not require *de novo* protein biosynthesis. Nonetheless, composition and mRNA dynamics of these temporal gene clusters may differ upon short and prolonged ERK activity, respectively. So far, temporal profiling of IEGs and DEGs has been limited to growth factor–induced ERK signalling. The described time course data obtained from HEK293ΔRAF1:ER now allows profiling of IEGs and DEGs both in response to short and prolonged ERK signalling and uncoupled from additional signalling pathways that would otherwise be activated in response to growth factor treatment.

Upon constant exposure to 4OHT (ON scenario, fig. 2.4) a total of 253 target genes were significantly induced and a total of 234 genes were significantly downregulated in HEK293ΔRAF1:ER cells. Remarkably, the majority of responding genes either steadily increased (189 of 253 upregulated genes) or steadily decreased (143 of 234 downregulated genes) upon sustained 4OHT exposure. In contrast, subsequent inactivation with U0126 two hours post induction resulted in transient up- or downregulation of these target genes (ON/OFF scenario) suggesting a direct link between ERK signalling input dynamics and mRNA response output dynamics. Overall, $\log_2$ fold changes for the most significantly up- and downregulated genes ranged between -1.4 for zinc finger and BTB domain containing protein 38 (*ZBTB38*) and +5.5 for tissue factor pathway inhibitor 2 (*TFPI2*). To identify primary response genes, CYHX was applied in parallel to 4OHT treatment. Among the 189 monotonically upregulated genes, 102 genes were still significantly induced upon parallel CYHX treatment (ON/CYHX scenario) and considered primary response genes (PRGs). A multitude of different mRNA dynamics was observed among PRGs with immediate, delayed and late responses. As elaborated above, these transcriptional waves have been termed immediate-early genes (IEGs) and delayed-early genes (DEGs). So far, classification of IEGs and DEGs has only been based on peak expression time points (Tullai et al. 2007; Amit et al. 2007). In this study, distinction of IEGs and DEGs is based on mathematical modelling (cf. sec. 3.2). In contrast to previous approaches this method allows precise quantification of transcriptional delays and to distinguish temporal gene clusters in a sustained signalling scenario where peak expression cannot be defined.

2.5 Mathematical modelling of mRNA dynamics

The mathematical model employed in this study was based on a minimal model of gene expression (Monod, Pappenheimer Jr., and Cohen-Bazire 1952; Gorini and Maas 1957; Berlin and Schimke 1965; Hargrove, Hulsey, and Beale 1991) with basal transcription rate $k_0 \in \mathbb{R}^{>0}$ and degradation rate $\gamma \in \mathbb{R}^{>0}$:

$$\xrightarrow{k_0} [\text{mRNA}] \xrightarrow{\gamma} \tag{2.1}$$

Assuming zero-order synthesis and first-order decay, a linear ordinary differential equation for this minimal model was stated as follows:

$$\frac{d\,[\text{mRNA}]\,(t)}{dt} = k_0 - \gamma\,[\text{mRNA}]\,(t) \tag{2.2}$$

As discussed in the introduction of this chapter, this minimal model serves as a reasonable approximation to study mRNA synthesis and degradation. However, the minimal model cannot account for dynamic changes in transcription as it assumes a constant rate of transcription. Furthermore, transcription occurs immediately and cannot be delayed. Hence, two additional parameters were introduced to overcome these limitations. An ERK activity-dependent parameter $k \in \mathbb{R}^{>0}$ was added to account for signalling-dependent transcriptional activity on top of basal transcriptional activity k_0 and a delay parameter $\Delta t \in \mathbb{R}^{\geq 0}$ was added to account for all steps that need to take place before ERK signalling–mediated transcription can start. These might include chromatin remodelling, transcription factor recruitment, and polymerase recruitment. The differential equation for the extended minimal model was stated as follows:

$$\frac{d\,[\text{mRNA}]\,(t)}{dt} = k_0 + k \cdot \text{pERK}\,(t - \Delta t) - \gamma\,[\text{mRNA}]\,(t) \tag{2.3}$$

Next, a solution of the differential equation was determined. Here, the first term of the solution corresponds to the general solution of the homogeneous equation and the second term corresponds to a special solution of the inhomogeneous equation:

$$[\text{mRNA}]\,(t) = [\text{mRNA}]_0\, e^{-\gamma t} \;+\; e^{-\gamma t} \int_0^t e^{\gamma \xi}\,(k_0 + k \cdot \text{pERK}\,(\xi - \Delta t))\; d\xi \tag{2.4}$$

For all simulations and model fitting described in the following sections, the initial starting concentration $[\text{mRNA}]_0$ was set to $\frac{k_0}{\gamma}$. This constraint implies that the system is in steady state as long as ERK is inactive ($\text{pERK}(t) = 0$). If ERK is activated ($\text{pERK}(t) = 1$) the system transitions to a new steady state at $\frac{k_0+k}{\gamma}$. This state transition and the biological interpretation of the different model parameters is schematically depicted in figure 2.5. Since

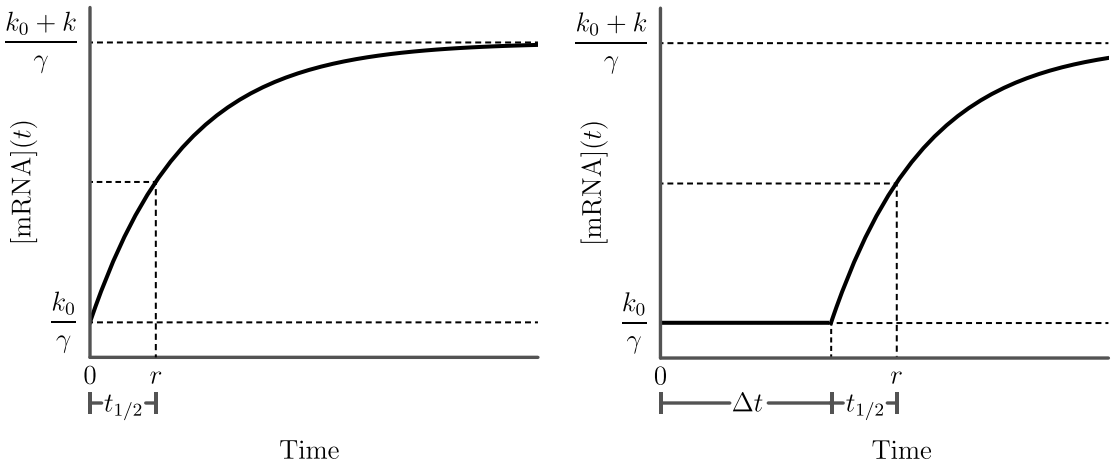


Figure 2.5 Schematic visualisation of model parameters.
The left panel depicts the state transition upon pERK induction from an initial mRNA concentration of $\frac{k_0}{\gamma}$ to a final mRNA concentration of $\frac{k_0+k}{\gamma}$ without any transcriptional delay ($\Delta t = 0$). Accordingly, the change in mRNA concentration upon pERK activation corresponds to $\frac{k_0+k}{\gamma} - \frac{k_0}{\gamma} = \frac{k}{\gamma}$. The response time r (which is defined as the point in time at which the modelled mRNA has reached half of its maximum induction) is equal to the sum of Δt and mRNA half-life $t_{1/2}$. Hence, for $\Delta t = 0$ it corresponds to $t_{1/2}$. For $\Delta t > 0$ (right panel) mRNA induction kinetics are shifted and r corresponds to $\Delta t + t_{1/2}$.

the model was also used to estimate mRNA half-life from induction kinetics, the relation between mRNA degradation rate and mRNA half-life needed to be specified. In general, the half-life $t_{1/2}$ of a product P is defined as the point in time by which the product's concentration has reached half of its initial concentration P_0. Mathematically, this relation can be stated as follows:

$$\frac{1}{2} \cdot P_0 = P\left(t_{1/2}\right) \tag{2.5}$$

The proposed framework, however, was build and constrained to model mRNA induction.[1] Thus, eq. 2.5 could not directly be used to establish the relation between mRNA degradation rate γ and mRNA half-life $t_{1/2}$ in this particular scenario. Yet, accumulation timing of an induced product is likewise related to its half-life. Hence, the relation between mRNA degradation rate and half-life was derived from induction kinetics instead. Here, it was

[1] Although, the framework could easily be adjusted to model mRNA decay or repression, for example by negating the step input function or by setting an arbitrary starting concentration of $[\text{mRNA}]_0 > 0$ and zeroing out k_0 and k.

assumed that the half maximum of mRNA induction is reached at $t = t_{1/2}$:

$$\frac{1}{2}\left(\frac{k_0}{\gamma} + \frac{k_0 + k}{\gamma}\right) = [mRNA]\,(t_{1/2}) \tag{2.6}$$

Next, eq. 2.4 was solved for $\text{pERK}(t) = 1$, $[\text{mRNA}]_0 = \frac{k_0}{\gamma}$ and $\Delta t = 0$, and equated with the half maximum induction level (eq. 2.6) to determine the relation between mRNA half-life $t_{1/2}$ and mRNA degradation rate γ:

$$\begin{aligned}
& \frac{1}{2}\left(\frac{k_0}{\gamma} + \frac{k_0 + k}{\gamma}\right) = [mRNA]\,(t_{1/2}) \\
\Leftrightarrow \quad & \frac{2k_0 + k}{2\gamma} = \frac{k_0}{\gamma} e^{-\gamma t_{1/2}} \; + \; e^{-\gamma t_{1/2}} \int_0^{t_{1/2}} e^{\gamma \xi}\,(k_0 + k)\; d\xi \\
\Leftrightarrow \quad & \frac{2k_0 + k}{2\gamma} = \frac{k_0 + k}{\gamma} \; - \; \frac{k}{\gamma} e^{-\gamma t_{1/2}} \\
\Leftrightarrow \quad & t_{1/2} = \frac{ln(2)}{\gamma}
\end{aligned} \tag{2.7}$$

Before the proposed mathematical framework was used for mRNA half-life estimations and identification of IEGs and DEGs in HEK293ΔRAF1:ER cells (ch. 3), its dynamic properties were explored in a series of *in silico* simulations.

2.6 *In silico* analysis predicts mRNA half-lives and transcriptional delays shape mRNA dynamics

Previous studies have shown that different signalling dynamics elicit a range of different mRNA dynamics (Saeki et al. 2009; Toettcher, Weiner, and Lim 2013; Zhang et al. 2014; Offermann et al. 2016; Wilson et al. 2017). But mRNA dynamics are not solely determined by signalling inputs. It is different combinations of mRNA half-lives and transcriptional delays that enlarge the number of possible gene expression profiles. In this section, the role of both mRNA half-lives and transcriptional delays for mRNA dynamics will be discussed in more detail.

As mentioned in the introduction to this chapter, the importance of mRNA half-life for the kinetics of gene induction was even acknowledged in the *lac operon* paper (Jacob and Monod 1961). Mathematical modelling of gene product synthesis and degradation provided a theoretical foundation for this assumption (Monod, Pappenheimer Jr., and Cohen-Bazire 1952; Gorini and Maas 1957; Yang et al. 2003). More recently, a variety of high-throuhput experiments have provided evidence to confirm this hypothesis (Shalem et al. 2008; Hao and Baltimore 2009; Elkon et al. 2010; Nagashima et al. 2015; Porter, Fisher, and Batchelor 2016; Cheng et al. 2017). Most importantly, the data confirmed that short-lived transcripts can

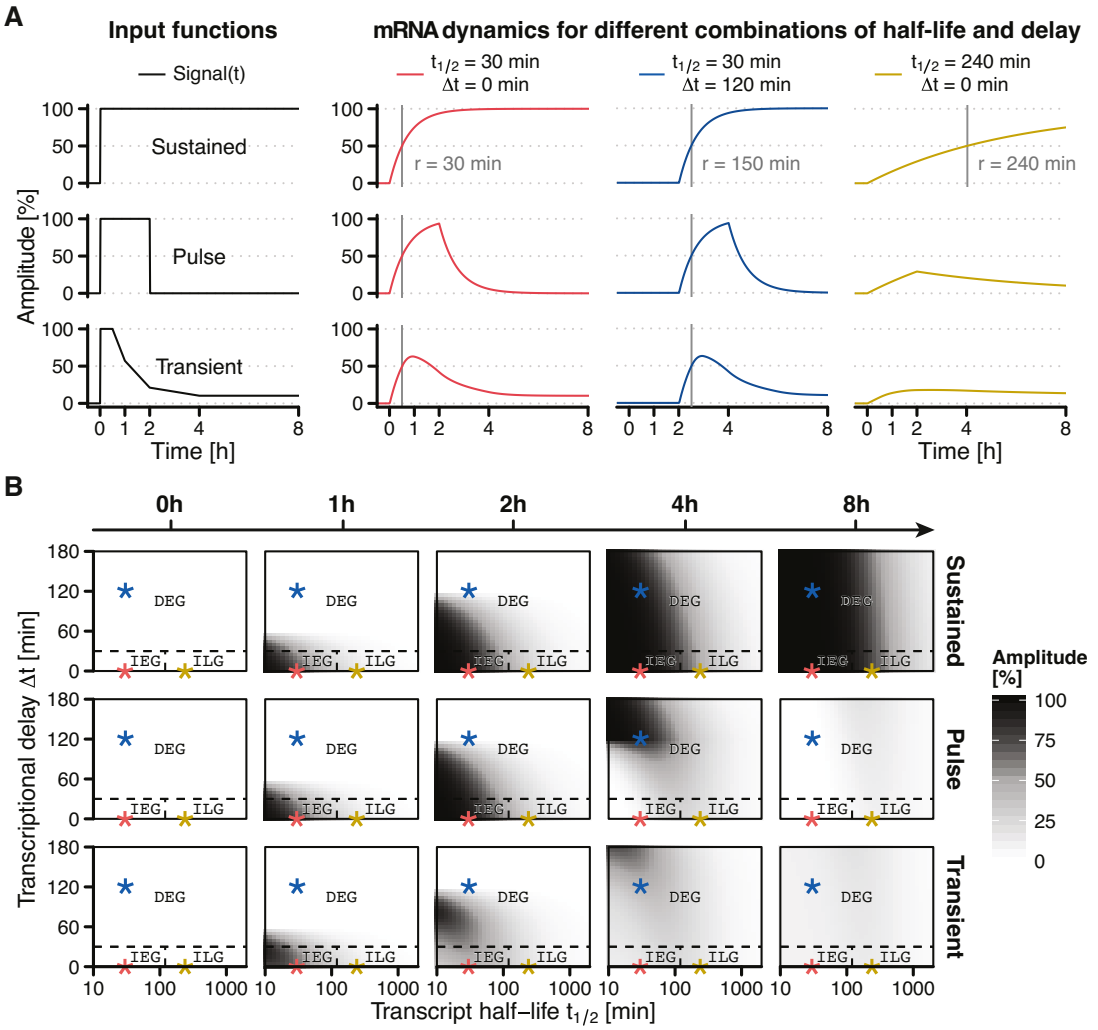


Figure 2.6 Simulation of primary response gene dynamics upon different signalling durations.

A: Different activation patterns of signalling molecules (input functions, left) can elicit multiple different response profiles (right) with different response times (r) depending on mRNA half-lives ($t_{1/2}$) and transcriptional delays (Δt). Rapid induction requires short half-lives (red lines). Late induction can be caused by transcriptional delays (blue lines), long half-lives (yellow lines), or combinations thereof. Decoding of signal duration depends on mRNA half-life. Short-lived mRNAs relay signal duration to response duration, whereas long-lived mRNAs decode signal duration to response amplitude (yellow lines).

B: Response amplitude for all simulated combinations of mRNA half-life and transcriptional delay. Response amplitude is shown over time (columns) and in respect to input function (rows). For sustained signalling, all primary response genes exceed their half maximum response amplitude. Pulse and transient signalling inputs are only sufficient for immediate-early genes and short-lived delayed-early genes. Long-lived mRNAs with half-lives greater 120 minutes require sustained signalling inputs to exceed their half maximum response amplitude. Example parameter sets displayed in (A) are marked with asterisk in (B). Dashed lines indicate cluster borders: IEG: Immediate-early genes, $t_{1/2} \leq 120$min and $\Delta t \leq 30$min. ILG: Immediate-late genes, $t_{1/2} > 120$min and $\Delta t \leq 30$min. DEG: Delayed-early genes, $\Delta t > 30$min.

be induced more rapidly than long-lived transcripts. This has important consequences for their potential response to different signalling dynamics. It implies that transient signalling inputs are sufficient for induction of short-lived mRNAs, whereas long-lived mRNAs require sustained signalling inputs to be able to reach their maximum response amplitude. In light of this important role of mRNA half-lives it seems inadequate that, so far, discrimination of PRGs responding to ERK signalling was only done according to the absence or presence of transcriptional delays in IEGs and DEGs, respectively (Amit et al. 2007; Tullai et al. 2007). An alternative classification of PRGs which accounts for both transcriptional delays and mRNA half-lives might thus more accurately reflect the different gene regulatory strategies observed in response to ERK signalling.

In this study, a new classification of PRGs is proposed which was initially based on a series of *in silico* simulations (fig. 2.6). Gene expression profiles were systematically predicted for a set of different parameter combinations in response to different ERK signalling scenarios. Also, the simulations were used to validate that the proposed mathematical framework can later be employed to model the obtained gene expression time-course data. First, different time-dependent input functions for $pERK\,(t)$ were incorporated into the model to simulate the effects of sustained, pulsed or transient ERK signalling on gene expression output dynamics (fig. 2.6A). Here, a unit step function was used to simulate sustained signalling :

$$pERK\,(t) = \begin{cases} 0 & \text{for} \quad t \leq 0 \\ 1 & \text{for} \quad t > 0 \end{cases} \tag{2.8}$$

A rectangular function with pulse duration of $p = 120\text{min}$ was used to simulate a pulsed signalling input:

$$pERK\,(t) = \begin{cases} 0 & \text{for} \quad t \leq 0 \\ 1 & \text{for} \quad t > 0 \ \wedge \ t \leq p \\ 0 & \text{for} \quad t > p \end{cases} \tag{2.9}$$

Lastly, a transient input function was generated to reflect a less synthetic but more physiological signalling input for pERK. The transient input function was based on linear interpolations of experimental data obtained from EGF-stimulated HEK293ΔRAF1:ER cells and its inference is described in section 3.3.

Next, three example gene transcripts were defined to simulate the response profiles for different temporal gene clusters (fig. 2.6A). A paradigmatic IEG was simulated with a mRNA half-life of 30 minutes, no transcriptional delay and a resulting response time of 30 minutes. A paradigmatic DEG was simulated with a mRNA half-life of 30 minutes, a transcriptional delay of 120 minutes and a resulting response time of 150 minutes. Lastly, an additional

example transcript was simulated with a mRNA half-life of 240 minutes, no transcriptional delay and a resulting response time of 240 minutes. The latter gene hereby serves as an example for a late response that can solely be attributed to mRNA longevity but not to transcriptional delays. In other words, this gene is immediately induced, but responds late and can thus be termed an immediate-late gene (ILG). For all genes, the response time corresponds to the time point when the respective gene reaches its half maximum amplitude. Since the response time is equal to the sum of a gene's transcriptional delay and its mRNA half-life, one can state that these two parameters in combination shape mRNA dynamics. Accordingly, the model confirmed that there are different gene regulatory strategies to achieve a late response in gene expression. Genes can respond late due to transcriptional delays, due to long mRNA half-lives or they can respond late by combining both.

When the three different example genes were compared across the three simulated signalling inputs, an additional implication of mRNA longevity became apparent. The immediate-early gene (characterised by its short half-life and no transcriptional delay) consistently reached at least 50% of its response amplitude across all simulated input scenarios (sustained, pulse, transient). Likewise, simulations predicted that the short-lived delayed-early gene is also capable of exceeding this level of response amplitude, but in a delayed fashion. Lastly, simulations demonstrated that only the long-lived immediate-late gene requires sustained signalling to reach at least 50 % of its response amplitude. An exploration of the parameter space systematically showed that these observations can generally be confirmed for IEGs, DEGs and ILGs (fig. 2.6B). For this, a total of about 250,000 parameter combinations were simulated for each of the three different input functions with transcript half-lives ranging between ten minutes and 24 hours and transcriptional delays ranging between zero and 180 minutes. Again, IEGs and short-lived DEGs responded across all simulated signalling inputs, but only ILGs (and long-lived DEGs) required sustained signalling to reach at least 50% of their response amplitude and were capable of translating different signal durations into different response amplitudes. It was hence concluded from simulations that long mRNA half-lives can govern duration-to-amplitude decoding.

It is important to note that response amplitudes are presented as relative values normalised to steady-state expression (cf. sec. A.5). Such normalised values ease the comparison of the timing between different genes during their transition from one steady-state to another. At the same time however, this representation cannot reflect absolute changes in mRNA concentration. Hence, throughout my thesis relative changes in expression (noted as amplitude [%]) are presented when the relation between mRNA half-life and signal duration decoding is described and absolute changes in expression (noted as $\log_2$ fold change) are presented, when quantitative aspects of mRNA expression are discussed. Still, all analyses of induction

kinetics presented here focus on exploring (relative or absolute) changes in gene expression. Absolute values of mRNA concentration are not provided as transcription and degradation rates share an arbitrary unit.

Altogether, in this chapter, different experimental model systems were presented to study gene-specific cellular responses. The highly controllable model system of HEK293ΔRAF1:ER cells was introduced in particular and the gene regulatory response to sustained and transient ERK signalling was profiled. The variety of observed mRNA dynamics called for introducing a mathematical framework to gain a better understanding of the parametric properties of these dynamics. Simulations based on this framework confirmed mRNA half-life as an important feature of mRNA dynamics. Two major implications of mRNA half-life were demonstrated: First, short mRNA half-lives enable the rapid induction of transcripts, whereas long mRNA half-lives enable the late response of transcripts. Secondly, short mRNA half-lives cause a relay of signal duration to response duration, whereas long mRNA half-lives enable duration-to-amplitude decoding. In the next two chapters, these *in silico* predictions will be tested by applying the introduced mathematical framework to the described gene expression time-course data and by systematically assessing mRNA half-life values and their role for gene expression timing and signal duration decoding.

3 Identification and characterisation of gene expression modules responding to ERK signalling

> *The structural message must be carried by a very short-lived intermediate both rapidly formed and rapidly destroyed during the process of information transfer. This is required by the kinetics of induction.*
>
> —François Jacob and Jacques Monod (1961)

3.1 Introduction

Only one year after the identification of the messenger RNA (Jacob and Monod 1961), its proposed short lifespan was experimentally validated in bacteria (Levinthal, Keynan, and Higa 1962) and, another year later, also in eukaryotic cells (Penman et al. 1963). Both studies used an antibiotic agent (Actinomycin D) that would specifically block biosynthesis of mRNA. An average mRNA half-life of two minutes in *B. subtilus* and of three to four hours in HeLa cells was estimated by chasing the degradation of the pre-existing mRNA fraction. Together, these findings confirmed Jacob and Monod's hypothesis that the kinetics of gene induction would require the mRNA to be a short-lived intermediate.

Whereas the importance of mRNA half-life for gene induction has been generally acknowledged, its role in shaping the transcriptional response to ERK signalling has not yet been studied in a genome-wide manner. Moreover, identification of ERK-regulated gene expression modules was only based on peak expression time points, but has not been related to any other dynamic property. In this chapter, the mathematical framework introduced in the previous chapter will be used to rank and classify the 102 identified PRGs responding to sustained ERK signalling in HEK293ΔRAF1:ER cells. Ranking and classification will be based on their dynamic properties, i.e. according to their response times, transcriptional delays and mRNA half-lives. At the end of the chapter, mRNA half-lives will be investigated more closely using a set of independent experimental approaches to take account for their immense importance in shaping gene expression dynamics.

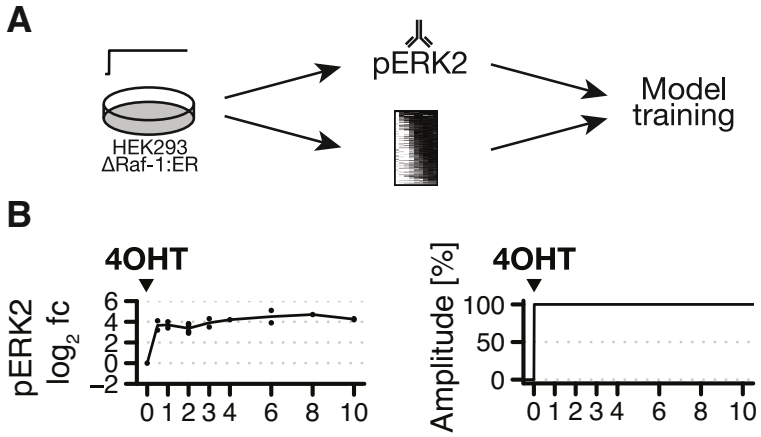


Figure 3.1 Model training scheme and signalling input validation.
A: HEK293ΔRAF1:ER cells were treated with 4OHT for constitutive induction of ERK signalling. Phosphorylation levels of ERK2 were measured with bead-based ELISA (Bio-Plex). mRNA time course expression data was measured using microarrays. This data served as a basis to train gene-wise model parameters.
B: pERK2 $\log_2$ fold change upon sustained activation (left) and deduced input function (right) used for model fitting. Average pERK2 $\log_2$ fold change upon 4OHT treatment equals 100% signalling amplitude.

With regard to gene expression dynamics, one must not forget to stress that mRNA half-lives and transcriptional delays are continuously distributed over genes. Hence, any grouping of genes into distinct clusters should be viewed as an indispensable heuristic in order to enable identification of functional differences across gene expression modules. In previous research, such abstractions helped to identify the module of delayed-early genes (DEGs) as a functional module of negative feedback regulators which control activity of immediate-early genes (IEGs) (Amit et al. 2007; Tullai et al. 2007), and, in this thesis, are used to identify and characterise a new gene expression module of immediate-late genes (ILGs).

3.2 Identification of IEGs, DEGs, and a new temporal cluster of immediate-late genes (ILGs)

Identification of distinct primary response gene clusters was based on the mathematical model of gene expression introduced in the last chapter. In particular, two specific features were used for cluster assignment. First, inferred transcriptional delays and goodness of fit estimates were used to discriminate between immediate and delayed primary response genes. Secondly, model-based mRNA half-life estimates were used to further discriminate between immediate-early and immediate-late primary response genes.

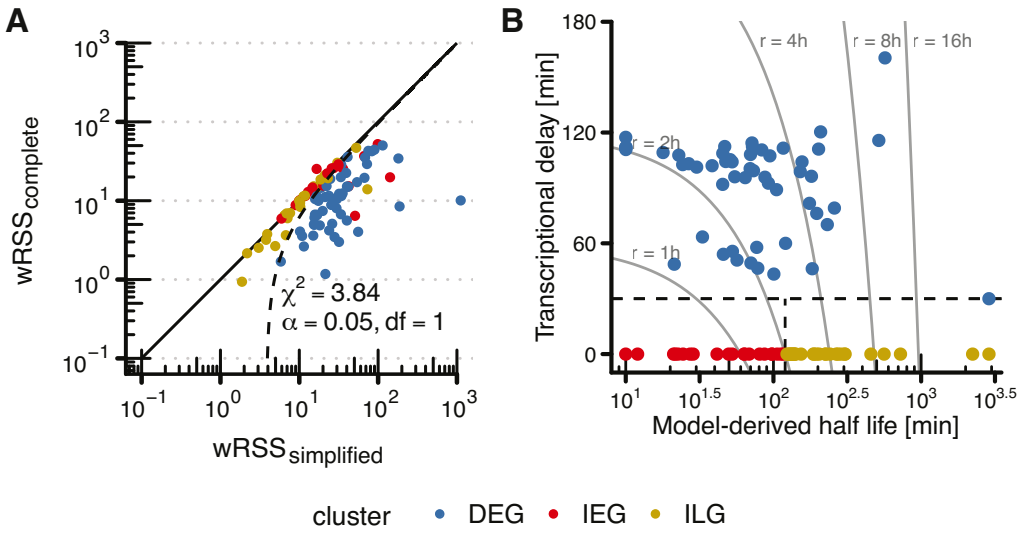


Figure 3.2 Cluster identification based on goodness of model fits and parameter estimates.
A: Sum of weighted squared residuals (wRSS) for simple (immediate) and complete (delayed) model reflect goodness of model fits across genes and were used to discriminate between delayed and immediate primary response genes. The complete model was rejected for genes with $\chi^2 < 3.84$ and $\Delta t < 30min$ to only accept significantly better fitted genes for the complete model and to reflect time intervals in sampling. Dashed line indicates chi-square threshold.
B: Parameter space for transcriptional delay and mRNA half-life allows visualisation of response times for IEGs, DEGs and ILGs. Genes on the same trajectory have the same response time, but response times are composed differently. For IEGs and ILGs, response times were solely determined by mRNA half-life, whereas response times of DEGs resulted from mixtures of mRNA half-life and transcriptional delay. Dashed lines indicate cluster borders.

To begin with, two different time-course measurements obtained from sustained ERK signalling were integrated and used as a training set for the mathematical framework. Stimulus-dependent phosphorylation of ERK was measured in a multiplex immunoassay (Bio-Plex) and gene expression data was obtained from Affymetrix Human Gene 1.0 ST microarray time course experiments (fig. 3.1A). Across all 4OHT treatment durations, pERK was reliably induced with mean $\log_2$ fold change of 3.89 ± 0.42 (fig. 3.1B). Based on this observation, a simplified step function for $pERK\,(t)$ was deduced (eq. 2.6) and incorporated as an input function into the extended model of gene expression (eq. 2.4).

Next, model parameters were fitted including basal transcription rate k_0, pERK-dependent transcription rate k, degradation rate γ and transcriptional delay Δt for each of the 102 significantly induced primary response genes identified in the previous chapter. In a simplified model transcriptional delay parameter Δt was left out and all remaining parameters were

fitted again. Afterwards, fits of the simplified and the complete model were compared. At this point one should note that, in general, more complex models are expected to yield more accurate fits, due to the additional degrees of freedom they provide. However, the increased accuracy comes at the cost of reduced generalisability, since more complex models might pick up variations that cannot be attributed to input dynamics, but are simply caused by biological or technical noise. Ultimately, favouring more complex models over a simple model with fewer parameters may result in overfitting (Abu-Mostafa, Magdon-Ismail, and Lin 2012, pp. 119). Hence, considerate model selection is key to come up with a framework which is not only able to explain the data used for model training, but which is also effective to predict and explain additionally acquired samples or data sets.

To ensure generalisability of the proposed models in the present case, a likelihood ratio test (Kreutz and Timmer 2009) was employed to assess differences between the sum of weighted squared residuals (wRSS) of the complete and simplified model (fig. 3.2A). As expected, the complete model generally yielded more accurate fits when compared to the simplified model, due to the additional degree of freedom provided by the transcriptional delay parameter. More precisely, complete model wRSS values were smaller for about three quarters of genes (78 in 102). Only for one gene (*DUSP1*) the complete model yielded a larger wRSS value and for 23 genes wRSS values were identical for both models. To account for the additional degree of freedom, the complete model was only accepted for genes with significantly enhanced fits (p-value < 0.05, chi-square $= 3.84$ for 1 degree of freedom). In total, 54 genes matched this criterion and, consequently, were identified as delayed-early genes (DEGs). For the remaining genes and for genes with $\Delta t < 30\ min$ (to reflect sampling intervals), the complete model was rejected and the simplified model was accepted instead. These genes were considered to be immediately induced upon ERK signalling and either identified as immediate-early genes or immediate-late genes, depending on whether their inferred mRNA half-life was smaller or greater than 120 minutes.

The scatter plot in fig. 3.2B shows the relation between inferred transcriptional delays and model-derived mRNA half-lives for all induced primary response genes. It visualises how genes were grouped into the different modules based on these two features and adopts the presentation of parameter spaces resulting from model simulations in the previous chapter (fig. 2.6B). In conclusion, estimated parameter values and goodness of fit statistics allowed subdivision of the 102 identified primary response genes into 21 rapidly induced IEGs, 54 late induced DEGs and 27 late induced ILGs. Classification into IEGs, DEGs and ILGs provided a coarse reflection of the temporal order of the ERK signalling–mediated gene response with IEGs responding first within two hours after ERK activation and both DEGs and ILGs responding subsequently with the vast majority of genes in these modules responding two

to eight hours after ERK activation. Interestingly, the parameter space plot (fig. 3.2B) already provides a comprehensive intuition about gene expression dynamics in response to ERK signalling without explicitly showing expression values, but only the inferred parameter values. In this representation, one can easily spot how the transcriptional response unfolds over time, starting with IEGs in the bottom left corner of the plot. DEGs then spread along both the x- and y-axis, as their kinetics are determined by both transcriptional delays and mRNA half-lives. On the contrary, ILGs spread only along the x-axis, as their kinetics are solely determined by mRNA half-lives.

Whereas such grouping of genes into temporal clusters eases comparison across literature, it tends to overshadow the aforementioned continuous nature of gene expression dynamics. To overcome this limitation, an unambiguous temporal ranking of all induced PRGs was established based on parameter estimates. For this, a response time r was defined for each gene which would correspond to the time when a particular gene reached its half maximum response amplitude. It could be calculated as the sum of transcriptional delays and model-derived half-lives ($r = \Delta t + t_{1/2}$). Calculated response times served two purposes. First, response time trajectories in fig. 3.2B demonstrated that certain DEGs and ILGs can respond at similar time points (about two to eight hours after ERK activation), but may use different regulatory strategies to achieve this late response. As elaborated above, they use different mixtures of transcriptional delays and mRNA half-lives. Secondly, by reducing the two temporal aspects of mRNA response into a single component, a precise temporal ranking of induced genes could be established. Based on this ranking, the early induction of IEGs and the late induction of DEGs and ILGs was recapitulated in a heat map representation of experimentally measured and computationally fitted expression values (fig. 3.3A). Here, expression values are presented as relative changes in amplitude to allow comparison of response dynamics across genes and a mean error (mean deviation of measured amplitudes and modelled amplitudes) is provided as a goodness of fit estimate. Overall accuracy of fits was very high with an average deviation of 6.8%. Only one gene (*NAGK*) stood out that had a markedly higher mean error of 50.0%, most probably caused by its non-monotonic induction kinetics that could not be captured by the model.

Response time compositions and accompanying box plots provided an overview of parameter distributions (fig. 3.3B). IEGs showed model-derived half-lives ranging from 10 (*DUSP1*) to 117 minutes (*EGR2*) and a short median response time of 53 minutes. DEGs showed a median model-derived half-live of 70 minutes, a median transcriptional delay of 102 minutes and a median response time of 160 minutes. Responding ILGs showed half-lives ranging from 124 minutes (*DUSP6*) up to 561 minutes (*QSOX1*) and a median response time of 204 minutes. For one DEG (*PPP1R15A*) and three ILGs (*BAIAP2*, *NR4A1*, *AKIRIN2*)

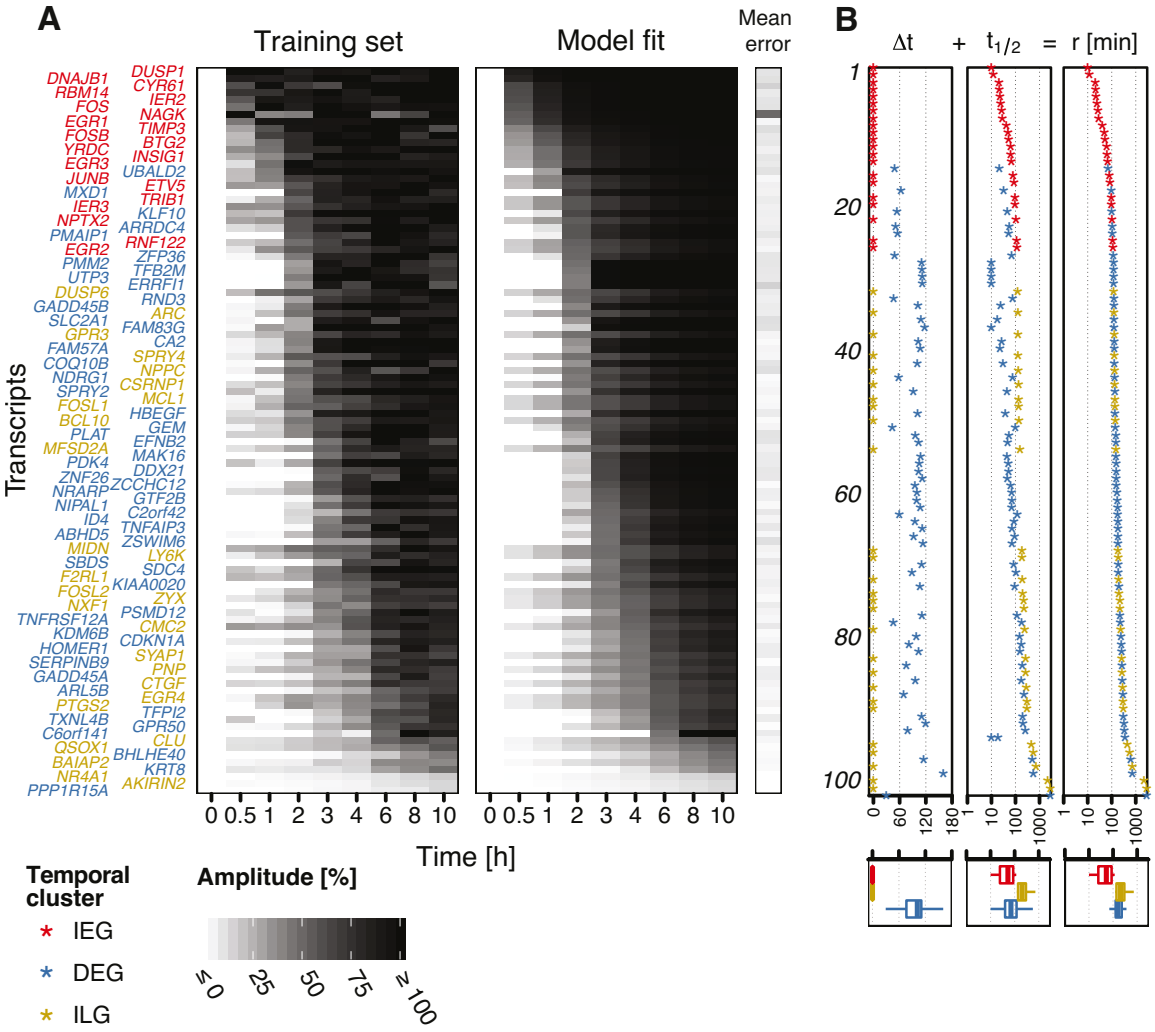


Figure 3.3 Model fitting and classification of primary response genes.
A: Measured gene expression kinetics and the resulting model fit of the gene expression model of all significantly induced primary response genes (FDR = 1 %). Gene expression is shown as percentage of response amplitude (cf. sec. A.5). Mean error was calculated as the average deviation from relative amplitudes serving as a goodness of fit measure. Gene names are alternately positioned.
B: Response time composition. For each gene, response time r is calculated as the sum of deduced transcriptional delay Δt and mRNA half-life $t_{1/2}$ and used for ranking. Upper panel shows individual parameter values for each gene. Lower panel shows boxplot distributions across temporal clusters. Gene-wise parameter estimates are listed in supplementary table B.3.

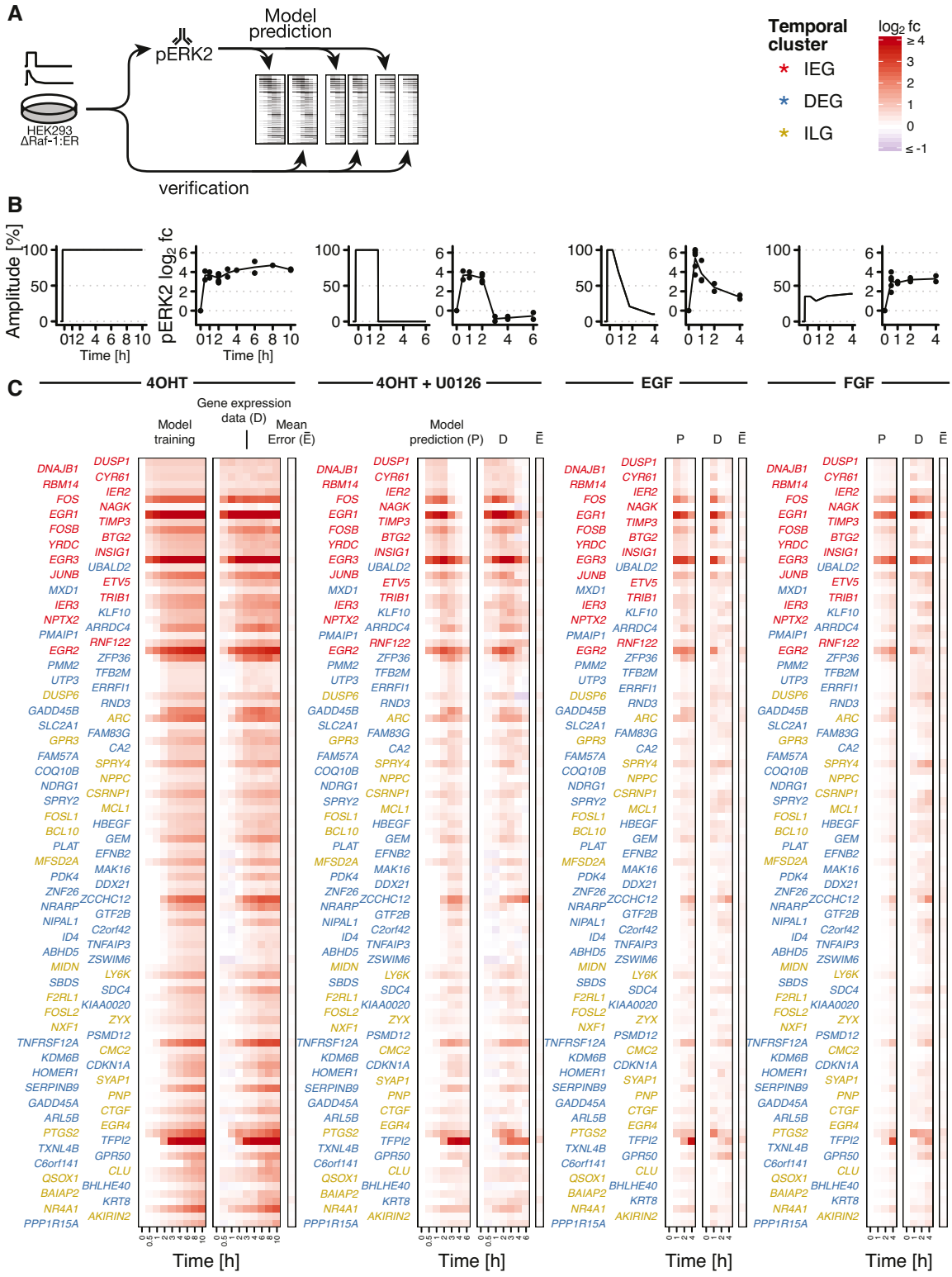
model-derived response times were greater than ten hours, the time span covered in the experiment. Again, all summarising values assigned to particular gene clusters like here need to be considered bearing in mind the continuous nature of gene expression parameters apparent in this analysis.

3.3 Model-derived parameters allow semi-quantitative prediction of mRNA log_2 fold changes

In a next step, it could be demonstrated that the parameter knowledge gained about PRGs induced upon sustained ERK signalling is generalisable and can be used to semi-quantitatively predict their behaviour upon different signalling scenarios (fig. 3.4). Sustained ERK signalling elicited by 4OHT was compared with a two-hour pulse (4OHT followed by U0126) and with EGF and fibroblast growth factor (FGF) treatment. Bead-based ELISAs confirmed that these stimuli indeed result in sustained, two-hour pulse and native growth factor–induced pERK dynamics, respectively (fig. 3.4B). EGF caused transient ERK activation (max. log_2 fold change: 5.43 ± 1.05) and FGF caused attenuated but sustained ERK activation (mean log_2 fold change: 2.83 ± 0.50) (fig. 3.4B). Growth-factor-mediated input functions required for prediction were generated from linear interpolations of pERK2 log_2 fold changes relative to mean induction in test condition. Deduced input functions were then incorporated into the established mathematical framework of gene expression (cf. eq. 2.4 and sec. A.5).

Strikingly, trained parameters allowed for semi-quantitative prediction of mRNA log_2 fold changes upon two-hour pulse signalling as well as upon growth factor treatments with EGF and FGF (fig. 3.4C). Although a wide range of fold changes was observed across responding genes and tested conditions, prediction errors were overall small. For the two-hour pulse experiment mean relative deviation from model predictions was $18.1\% \pm 11.8\%$. For growth factor treatments it was $19.1\% \pm 15.0\%$ for EGF and $20.5\% \pm 12.3\%$ for FGF respectively. The observation that predicted log_2 fold changes were slightly more accurate for the two-hour pulse experiment could be neglected, since the differences were only subtle and not significant. However, the tendency might be explained by the fact that the first four time points for the two-hour pulse (0, 0.5, 1 and 2 hours) were identical with the first four time points used in the training set time course. Only the last three time points which were acquired after application of the MEK inhibitor U0126 (3, 4 and 6 hours) were truly predicted in this case. Hence, the presented prediction error for the two-hour pulse time course must be considered carefully as it resulted from a mixture of model fitting errors and actual prediction errors.

A close inspection of predicted and measured log_2 fold changes further suggested that early (on-)kinetics were overall more accurately predicted than late (off-)kinetics (fig. 3.4C). Again, in case of the two-hour pulse experiment this observation can easily be explained by the mixture of training and test data. Especially for IEGs and DEGs the timing for off-kinetics was either over- or underestimated for a subset of genes, whereas errors for on-kinetics were smaller. In case of EGF treatment, predicted on-kinetics accurately matched measured log_2 fold changes, but off-kinetics were systematically underestimated. Especially for some IEGs


A
Model prediction
pERK2
HEK293 ΔRaf-1:ER
verification
Temporal cluster
IEG
DEG
ILG
log2 fc
≥ 4
3
2
1
0
≤ -1
B
Amplitude [%]
pERK2 log2 fc
Time [h]
C
4OHT
4OHT + U0126
EGF
FGF
Model training
Gene expression data (D)
Mean Error (Ē)
Model prediction (P)
P
D
Ē
Time [h]

measured off-kinetics were much faster than predicted ones. Here, incorporation of additional input features like transcriptional feedbacks involved in EGF-mediated ERK sigalling (Amit et al. 2007; Tullai et al. 2007) might improve model predictions, but most probably play only a minor role for output dynamics when compared to the major determinants identified here, namely ERK signalling input dynamics and gene-specific model parameters for mRNA degradation and transcriptional delay.

Cluster-wise prediction errors showed that fitting and prediction of IEG kinetics was slightly worse than fitting and prediction for genes from other temporal clusters (fig. 3.5). Interestingly, ILG kinetics were consistently better to fit and predict across signalling inputs when compared to IEGs and DEGs. When comparing fitting and prediction errors, one could see that not only median errors were smaller for fits, but also interquartile ranges were much smaller, which matched general expectations for model fitting.

In conclusion, the overall predictive power of the mathematical framework implicated that the proposed model is generalisable and that a sparse set of extrinsic and intrinsic features is sufficient to explain observed gene response output dynamics. First, ERK signalling input dynamics serve as a main extrinsic force controlling induction and repression of gene response. Secondly, the plethora of different on- and off-kinetics results from gene-specific, intrinsic properties, namely transcriptional delays and mRNA half-lives.

Figure 3.4 ***(facing page)*** **Semi-quantitative prediction of mRNA log_2 fold changes upon different signalling scenarios.**
A: Gene expression upon different stimulations was predicted based on fitted model parameters (cf. sec. 3.2) and measured pERK2 levels (shown in **B**). Predictions were verified with gene expression time-course data.
B: Signalling input conditions for different signalling scenarios (for each panel, left side shows deduced input function and right side empirical pERK2 measurements): Sustained ERK signalling (4OHT), two-hour pulse ERK signalling (4OHT+U0126), growth factor signalling (EGF: epidermal growth factor, FGF: fibroblast growth factor). Deduced input functions: 100% signalling amplitude corresponds to mean induction in training condition (4OHT). Growth factor-induced input functions are linear interpolations of pERK2 log_2 fold changes relative to mean induction in test condition.
C: Model predictions are verified with actual gene expression time course data. Heat maps show log_2 fold changes of 102 induced primary response genes. Text colour codes for temporal clusters. P: Model prediction. D: Gene expression data. E: Mean error = mean of log_2 fold change residuals.

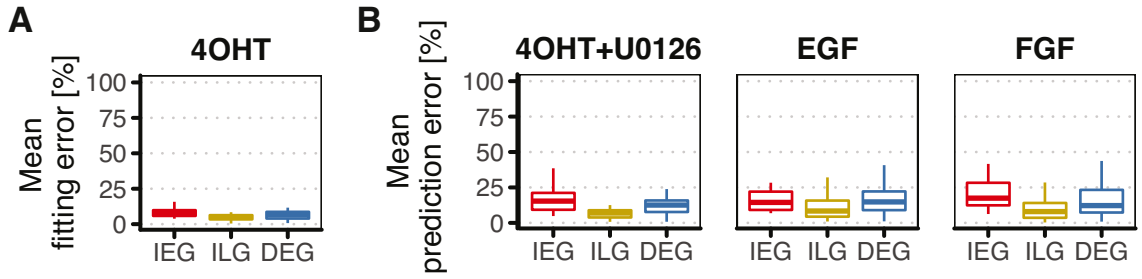


Figure 3.5 Goodness of gene expression predictions.
A: Mean fitting errors across temporal clusters, calculated as the mean of weighted residuals.
B: Mean prediction errors across temporal clusters for tested signalling scenarios, calculated as the mean of weighted residuals.
Data information: Throughout the thesis, boxplots show median and inter-quartile range (IQR) information. IQR is extended with whiskers to the largest and smallest value respectively, but no further than 1.5x IQR from hinges.

3.4 High-throughput measurements confirm short mRNA lifespan of IEGs and longevity of ILGs

Having demonstrated the applicability and predictive power of the mathematical framework for different ERK signalling scenarios in the previous section, the following two sections will focus on validating parameter estimates for mRNA half-lives and transcriptional delays respectively. Experimental validation of model-derived estimates is key to support the claim that these two dynamic properties indeed are intrinsic features of the induced genes and hence can be used to explain the observed gene regulatory dynamics.

As discussed in the introduction to this chapter, initial efforts to quantify mRNA half-lives made use of an experimental approach where transcription is blocked entirely and the degradation of transcripts is chased thereafter. However, since the blockage of mRNA synthesis was considered highly stressful for cells, alternative methods were developed to experimentally assess mRNA half-life in a less invasive way. Instead of directly chasing mRNA decay after transcriptional shutdown, modified ribonucleotides were introduced to cells to specifically label newly synthesised mRNA. Such metabolic labelling approaches allowed for estimation of mRNA half-life by comparing nascent (labelled) and pre-existing (unlabelled) mRNA levels (pulse approach) or by chasing the decay of a pulse-labelled mRNA fraction (pulse-chase approach). The first reported experiment involving a labelled ribonucleotide analogue to study mRNA half-life was performed in 1971 (Kramer and Hilz 1971). HeLa cells were cultured with a radioactive isoform of uridine ([^{3}H]uridine) and mRNA half-life was determined by measuring the decay of the radioactively labelled mRNA after change

to a [^{3}H]uridine-free culture medium. Similar experiments were also performed using different radioactive ribonucleotides such as [^{3}H]adenine, [^{3}H]guanosine or [^{14}C]adenine (Sheiness, Puckett, and Darnell 1975; Petersen, McLaughlin, and Nierlich 1976). In general, these experiments again confirmed a short lifespan for mRNA compared to other RNA species like ribosomal RNA (rRNA) and transfer RNA (tRNA).

Still, experimental approaches involving radioactive substances were neither regarded as truly non-invasive measurements, since DNA double strand breaks could be caused and disturb overall cellular metabolism. In response to that even less invasive methods were developed. Instead of radioactively labelled isoforms, thiol-linked ribonucleotides were used to label RNA molecules such as 4-thiouridine (4SU) and 6-thioguanosine (6SG) (Melvin et al. 1978). Whereas these substances were less harmful to both treated cells and the handling experimenter, the novel approach came at the cost of requiring a laborious purification protocol to separate labelled RNA from unlabelled RNA. In consequence, popularity of the approach was limited and when the advent of high-throughput methods for the first time enabled genome-wide assessments of RNA half-lives, laboratories preferred transcriptional shut-down approaches instead (Raghavan et al. 2002; Frevel et al. 2003; Yang et al. 2003; Raghavan and Bohjanen 2004).

Only recently, a simplified purification protocol (Dölken et al. 2008) led researchers to revisit thiol-based labelling methods after its potential to assess genome-wide estimates of RNA half-lives was first demonstrated in a high-troughput experiment (Cleary et al. 2005). It sparked a series of new investigations on RNA metabolism in general and mRNA half-life in particular. For example, new genome-wide assessments were made in yeast (Miller et al. 2011), murine cells (Friedel et al. 2009; Rabani et al. 2011; Schwanhäusser et al. 2011; Marzi et al. 2016) and human cells (Friedel et al. 2009; Marzi et al. 2016). Over the years, improved or modified protocols were published (Neymotin, Athanasiadou, and Gresham 2014; Schwalb et al. 2016; Herzog et al. 2017; Schofield et al. 2018) and a range of computational frameworks was developed to aid interpretation of these experiments (Schwalb et al. 2012; Rabani et al. 2014; Pretis et al. 2015; Jürges, Dölken, and Erhard 2018). Among the newly developed protocols, two stand out in particular, as they now even allow for discrimination of nascent and pre-existing mRNA without the need for biochemical fractionation. These protocols demonstrate that 4SU can be converted into cytosine during reverse transcriptase reaction either by alkylation (Herzog et al. 2017) or by oxidative nucleophilic aromatic substitution (Schofield et al. 2018). When analysed in an RNA-sequencing experiment, the conversions appear as T-to-C mutations and allow for *in silico* separation of mRNA fractions. Again, the ratio of computationally inferred mRNA fractions then serves as a proxy to compute mRNA half-lives (Jürges, Dölken, and Erhard 2018).

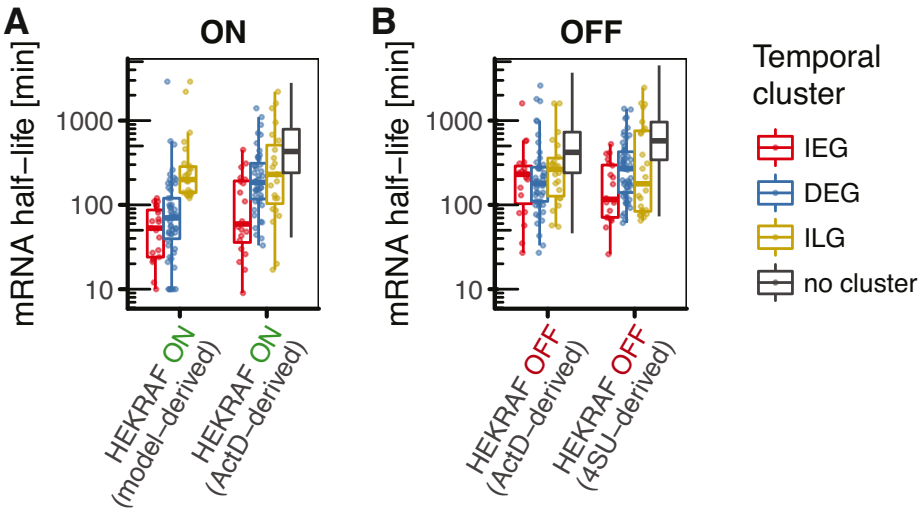


Figure 3.6 Immediate-late genes (ILGs) have long mRNA half-lives.
A: Boxplot comparison of mRNA half-life estimates based on modelling of gene induction (model-derived) and transcriptional shutdown experiments (ActD-derived) from 4OHT-pretreated cells.
B: Boxplot comparison of ActD-derived mRNA half-life estimates and metabolic labelling (4SU-derived) taken from unstimulated cells.
Estimates from 4OHT-pretreated HEK293ΔRAF1:ER cells (ON panel) are more appropriate to characterise induced genes than estimates from unstimulated cells (OFF panel). Genes not assigned to any cluster are shown in grey.

Most of the methods described above have recently been reviewed in the literature (Tani and Akimitsu 2014; Wada and Becskei 2017) and can generally be assigned to one of three different categories: transcriptional shutdown, metabolic labelling and gene induction. In my thesis, initial mRNA half-life estimates were based on a gene induction approach, as a subset of genes responding to ERK signalling was specifically induced and half-lives were inferred from response dynamics. In a second step, initial estimates were now validated using both transcriptional shutdown experiments (using ActD) and metabolic labelling experiments (using 4SU) (fig. 3.6, cf. fig. 2.3 for sampling and sec. A.6 for derivation). ActD-derived half-lives were determined in both 4OHT-pretreated and untreated HEK293ΔRAF1:ER cells (fig. 3.6A+B). On the contrary, 4SU-derived half-lives were only determined in untreated HEK293ΔRAF1:ER cells (fig. 3.6B), since this approach assumes steady-state gene expression.

ActD-derived mRNA half-life estimates from 4OHT-pretreated HEK293ΔRAF1:ER cells confirmed short mRNA half-lives for IEGs (median = 60min) and longer mRNA half-lives for ILGs (median = 230min, fig. 3.6A). Whereas median values were very similar to gene induction–based estimates (53min for IEGs and 204min for ILGs), variation was larger for ActD-based estimates. In contrast to pretreated cells, data from uninduced cells only par-

tially recapitulated half-life estimates (fig. 3.6B). On the one hand, longevity of ILGs could again be confirmed since estimates from uninduced cells were in a similar range (median = 265min for ActD-derived, median = 181min for 4SU-derived). On the other hand, however, estimates for IEGs were systematically larger in uninduced cells when compared to 4OHT-pretreated cells (median half-life of 53min and 60min in pretreated cells and median half-life of 116min and 230min in uninduced cells). Possible causes for these differences could be of biological or technical origin. For example, an active destabilisation of IEGs in response to ERK signalling has been reported and attributed to gene products of some DEGs (Avraham and Yarden 2011) and could potentially account for shorter half-lives of IEGs in 4OHT-pretreated cells. From a technical perspective, it is more difficult to accurately determine half-lives for lowly expressed genes, especially when expression values were measured with microarrays, which is the case for the presented Actinomycin D data sets. Although genes responding to ERK signalling do show basal transcriptional activity in absence of active ERK, technical noise in a microarray experiment potentially obscures baseline expression and may lead to erroneous half-life estimates.

Whereas longevity of ILGs was confirmed across all exerted approaches, gene induction-based mRNA half-life estimates for DEGs could neither be recapitulated in transcriptional shutdown experiments nor in metabolic labelling experiments. Notably, for DEGs, estimates from validation experiments were in range of estimates for ILGs (> 120min), whereas estimates based on modelling of gene induction had suggested shorter half-lives for DEGs in range of IEGs (< 120min). This discrepancy could be attributed to an oversimplified model for transcriptional delays. Despite its demonstrated predictive power across temporal clusters (fig. 3.4), the mathematical framework is potentially more appropriate for immediately induced genes (IEGs and ILGs) than for delayed genes (DEGs), as it assumes an instant change of the rate constant for transcription once ERK is activated and promoters are accessible (step function). This assumption is very well justified for poised promoters of immediately induced genes. For delayed genes, changes in transcription rate should nearly correspond to a (delayed) step function as well, since changes in chromatin accessibility of DEG promoters follow a multi-step process and, in consequence, are highly non-linear. In turn, changes in transcription are expected to be subtle at the beginning, and very steep at the end, once the chromatin is fully accessible. Still, these differences may account for a systematic underestimation of mRNA half-lives in delayed genes, as the continuous change of transcription rates is discretised in a delayed step function. It was therefore concluded that half-lives of IEGs and ILGs can certainly be estimated from gene induction kinetics, whereas half-lives of DEGs are more reliably determined in transcriptional shutdown and metabolic labelling experiments.

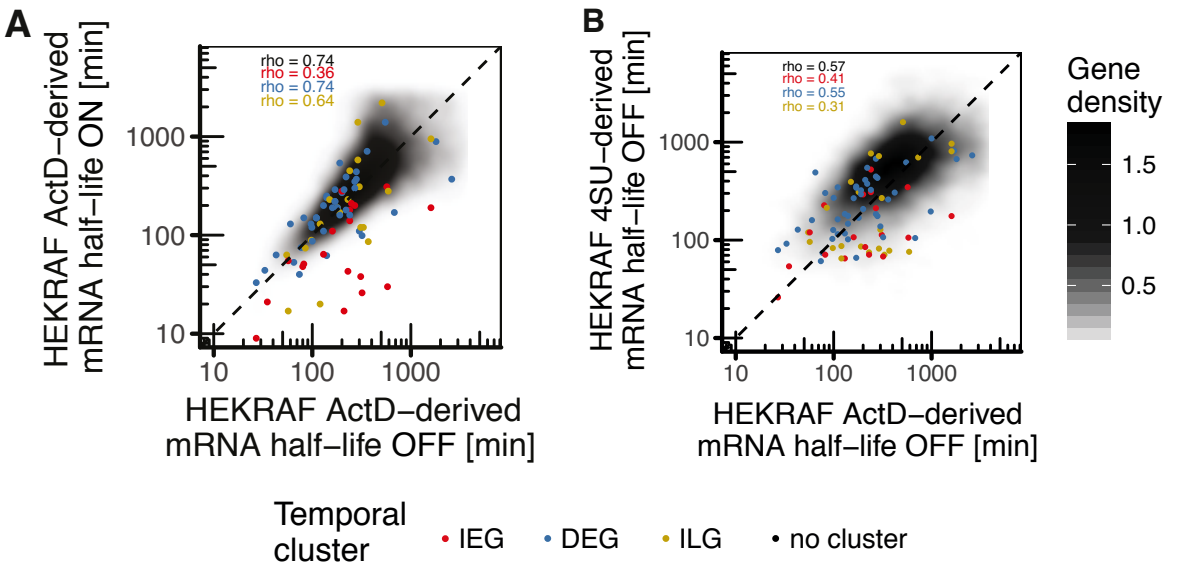


Figure 3.7 mRNA half-life comparisons for HEK293ΔRAF1:ER cells.
A: Treatment comparison: ActD-derived half-lives in 4OHT-pretreated versus untreated HEK293ΔRAF1:ER cells. Spearman correlation is indicated for all genes and for each gene cluster.
B: Method comparison: ActD-derived half-lives compared to 4SU-derived half-lives in untreated HEK293ΔRAF1:ER cells.

In summary, validation experiments confirmed short mRNA half-lives for IEGs and longevity of ILGs, but suggested longer half-lives for DEGs compared to estimates from mathematical modelling of gene induction kinetics. At this point, to avoid any misconception about the notion of short or long mRNA lifespans, half-lives of induced genes should be put into perspective to half-life estimates for all genes in the cell. In fact, half-lives for all three temporal clusters were shorter than median half-life estimates for genes not assigned to any temporal cluster. Here, genes not assigned to any PRG cluster include secondary response genes (87 genes), but mainly consist of uninduced genes not responding to ERK signalling (19299 genes). In consequence, the identified longevity of ILG mRNAs must be considered in relation to the extremely short mRNA lifespan of IEGs and must not be mistaken with a much longer mRNA lifespan observed in the majority of uninduced genes in general or certain housekeeping genes in particular.

When comparing mRNA half-life estimates based on a range of different experimental methods an obvious question arises: Are overall estimates reasonably similar across experimental methods and do some methods provide more accurate measurements of mRNA half-life than others? The initial paper on genome-wide 4SU-based measurements of mRNA half-life demonstrated that estimates indeed do correlate across methods in mammalian cells

(Dölken et al. 2008). However, comparisons of transcriptional shutdown and metabollic labelling methods in budding yeast (*S. cerevisiae*) revealed dramatic differences and poor correlations of mRNA half-life estimates (Sun et al. 2012). Poor correlation in yeast was attributed to extensive compensatory effects of transcriptional shutdown by global stabilization of the transcriptome (Sun et al. 2012). From these experiments, it was concluded that transcriptional shutdown experiments, although yielding reproducible half-life estimates across yeast strains and experimental conditions, are very disruptive and metabolic labelling experiments should be favoured to generate more accurate half-life estimates (Wada and Becskei 2017).

Although the reported poor correlation between transcriptional shutdown and metabolic labelling approaches in yeast is alarming, the data obtained from HEK293ΔRAF1:ER cells suggested that both methods provide reasonable estimates for both ERK signalling–induced and uninduced genes. Three different assessments were made to arrive at this statement. First, the same method (transcriptional shutdown) was compared across two different treatment conditions (untreated versus 4OHT-pretreated, fig. 3.7A). Secondly, two different methods (transcriptional shutdown versus metabolic labelling) were compared for the same experimental condition (untreated cells, fig. 3.7B). Thirdly, median mRNA half-life estimates for all genes were derived from the three different data sets and were compared to two different literature data sets, one of them based on a transcriptional shutdown approach (Yang et al. 2003) and one of them based on a metabolic labelling approach (Friedel et al. 2009, fig. 3.8).

The comparison of transcriptional shutdown experiments yielded overall high correlation. More precisely, ActD-derived mRNA half-lives in 4OHT-pretreated and untreated samples correlated well for DEGs and ILGs (fig. 3.7A, Spearman's rho of 0.74 and 0.64 respectively). Only IEG estimates showed rather weak correlation, hinting at potential destabilization of IEGs upon ERK signalling or technical difficulties in assessment, as discussed above. When considering all genes expressed, an overall high correlation was found (Spearman's rho = 0.74) for the two different experimental conditions. However, variability was lower for short-lived genes than for long-lived ones. This observation could be explained by the fact that transcriptional shutdown was only chased for up to eight hours. On this time scale, genes with half-lives greater than eight hours are difficult to quantify as mere halving of expression values could be obscured by noise. Hence, a longer chase could potentially improve reproducibility of estimates for long-lived genes and increase overall correlation.

Correlation between mRNA half-life estimates from the two different methods was lower than correlation between the two ActD-derived data sets, but still surprisingly high (fig. 3.7B, Spearman's rho = 0.57) when considering the alarmingly poor correlation across methods reported in yeast (Sun et al. 2012). No clear differences between short and long-lived genes

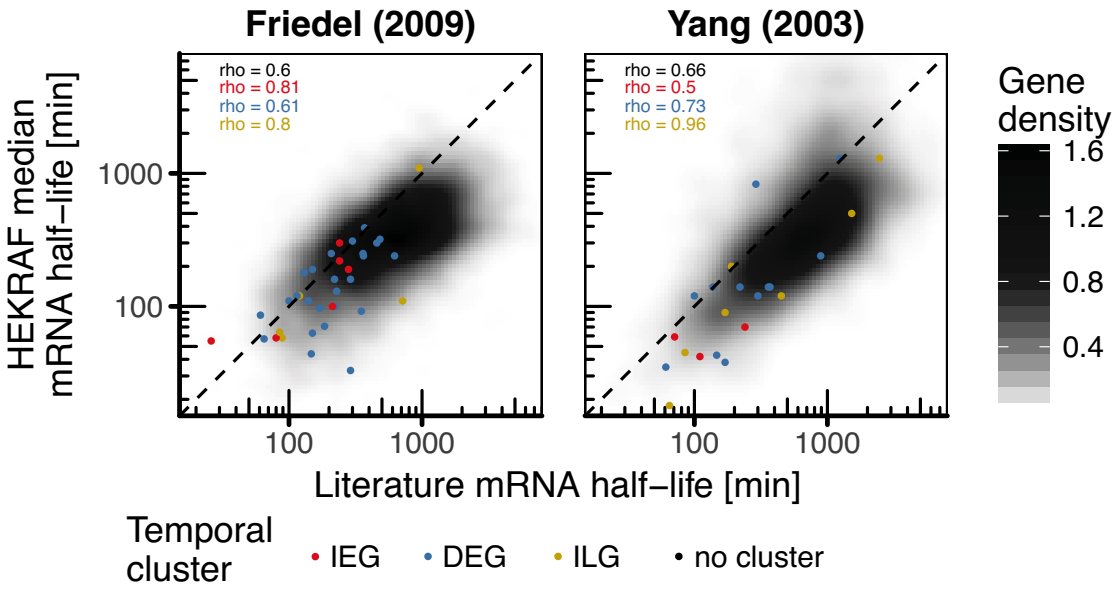


Figure 3.8 Comparison to literature mRNA half-lives.
HEK293ΔRAF1:ER median mRNA half-lives were compared to published mRNA half-lives in two different studies. Friedel et al. provided 4SU-based measurements from human B-cells and Yang et al. provided ActD-based measurements from human hepatocellular carcinoma cell line HepG2.

were detected. However, correlation of estimates for the three different temporal clusters was lower than for the overall data set (fig. 3.7B). Still, overall correlation between the two different approaches suggested to consider all data sets in combination to define a comprehensive list of mRNA half-lives in HEK293ΔRAF1:ER cells.

A final list for mRNA half-lives in HEK293ΔRAF1:ER cells was generated by taking the median half-life value from all obtained data sets for each gene. Gene-wise median mRNA half-life estimates were then tested against two human data sets on mRNA half-lives. Good correlation was found with both data from metabolic labelling experiments and transcriptional shutdown experiments (fig. 3.8, Spearman's rho = 0.60 with Friedel et al. 2009 and rho = 0.66 with Yang et al. 2003). Although median estimates correlated well with reported literature values, systematic differences were observed. Published 4SU-derived estimates were in a similar range for short-lived mRNAs, but were consistently larger for long-lived mRNAs. For published ActD-derived estimates, half-lives were consisently larger across the entire range of estimates. On average, half-life estimates were 370 minutes larger in published metabolic labelling data, corresponding to an average relative difference of 17.6%, and 207 minutes larger in transcriptional shutdown data, corresponding to an average relative difference of 16.5%. Systematic differences as such could be attributed to technical compli-

cations like partial sample degradation during preparation, or to difficulties in adequately normalizing the obtained data. However, since differences could likewise originate from cell-type specific differences in overall transcript stability, for instance potentially caused by high cellular ribonuclease activity, the comparative analysis performed here overall supported validity of mRNA half-life estimates and provided a catalogue of mRNA half-lives for the experimental model system used in this study (cf. supplementary table EV2 in the online version of the original manuscript, Uhlitz et al. 2017).

3.5 ILGs possess GC-rich promoters and are transcribed immediately

Mathematical modelling of PRGs suggested two main dynamic characteristics for ILGs. The previous sections determined and validated long mRNA half-lives for ILGs as one major characteristic. This section now investigates immediacy of mRNA synthesis across clusters, since modelling predicted that ILGs would respond late like DEGs, but would be induced immediately like IEGs. It has been previously reported that IEGs and DEGs differ in their promoter architecture (Tullai et al. 2007; Ramirez-Carrozzi et al. 2009; Avraham and Yarden 2011). Compared to the overall GC content in the human genome, which approximates to 41%, GC content in promoters is generally enriched and amounts to about 57%.[1] It was shown that GC enrichment is particularly strong in IEG promoters, allowing for instant activation independent of nucleosome remodellers. In contrast, promoters of genes responding with a delay commonly possess (relatively) GC-poor promoters, facilitating their dependence on remodellers and thereby delaying their induction (Ramirez-Carrozzi et al. 2009).

To calculate promoter GC content, stretches of ±1000bp relative to the transcription start site (TSS) of each primary response gene identified in this study were examined (fig. 3.9A). In accordance with the literature, it was found that IEGs possess GC-rich promoters with a median GC content of 60%. In contrast DEGs had a lower median GC content of 57% which precisely matched the median GC content of promoters in the human genome in general. Notably, promoter analysis revealed that ILGs, like IEGs, possess GC-rich promoters with a median GC content of 60% which is significantly higher (Wilcoxon rank sum test p-value = 1.4*10-4) than GC content determined for DEG promoters. In summary, examination of promoter GC content supported the conjecture that ILGs possess permissive promoters like IEGs that are potentially more accessible than DEG promoters. Hence, immediate induction of ILGs is potentially facilitated by their permissive promoter architecture.

Since promoter GC content only provided an indirect measure hinting at immediate in-

[1] Percentages were calculated for human reference genome GRCh37 and GENCODE annotation v19.

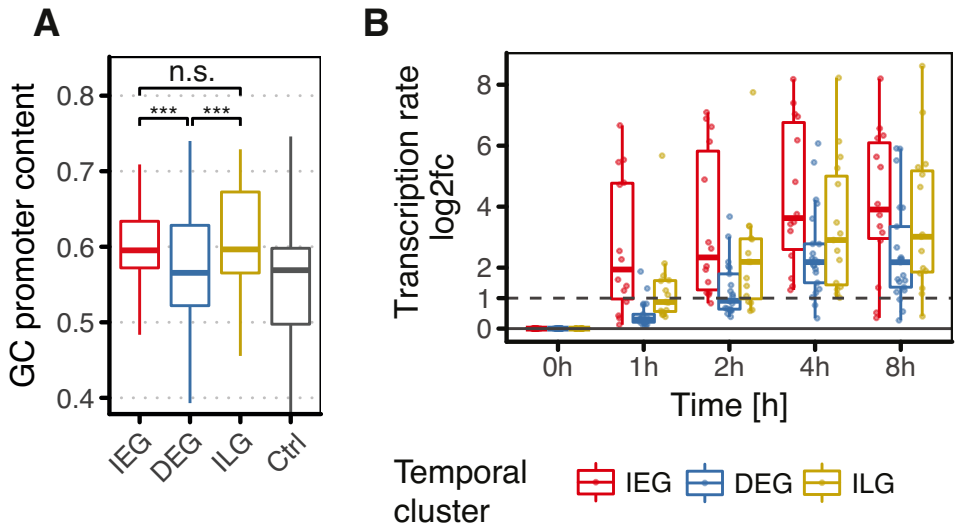


Figure 3.9 Immediate-late genes (ILGs) possess GC-rich promoters and are transcribed immediately.
A: Promoter GC content in IEGs, DEGs and ILGs and a control group (Ctrl) of 100 random promoters in the human genome. Promoters were defined as 2000bp stretches centred around transcription start sites (TSS) using the human reference genome assembly `hg19`. Wilcoxon rank sum was used to test for significant differences (n.s.: not significant. ***: P-value < 0.001).
B: Log_2 fold changes of transcription rate in 4OHT-treated HEK293ΔRAF1:ER cells derived from metabolic labelling (4SU) RNA-sequencing data document immediate transcription of IEGs and ILGs but delayed transcription of DEGs. Dashed line indicates doubling of transcription rate.

duction of ILGs, a direct experimental assessment of transcription rates was necessary for validation. For this, one-hour pulses of metabolic labelling with 4SU followed by RNA-sequencing were performed in 4OHT-pretreated HEK293ΔRAF1:ER cells (cf. fig .2.3B for sampling). As elaborated above, metabolic labelling data can be used to infer mRNA half-lives by considering the ratio between labelled and unlabelled mRNA. However, the data can likewise be used to directly assess transcription rates, since the labelled mRNA fraction corresponds to nascent mRNA transcribed after addition of the labelling reagent. Accordingly, changes in nascent mRNA levels after 4OHT treatment reflect changes in transcription rates in response to ERK signalling.

In agreement with the proposed model, the median transcription rate of ILGs was approximately doubled just one hour after induction of ERK signalling with 4OHT (fig. 3.9B), suggesting immediate transcription. DEGs in contrast required two hours of 4OHT treatment until median transcription rate was doubled, suggesting delayed transcription. Overall,

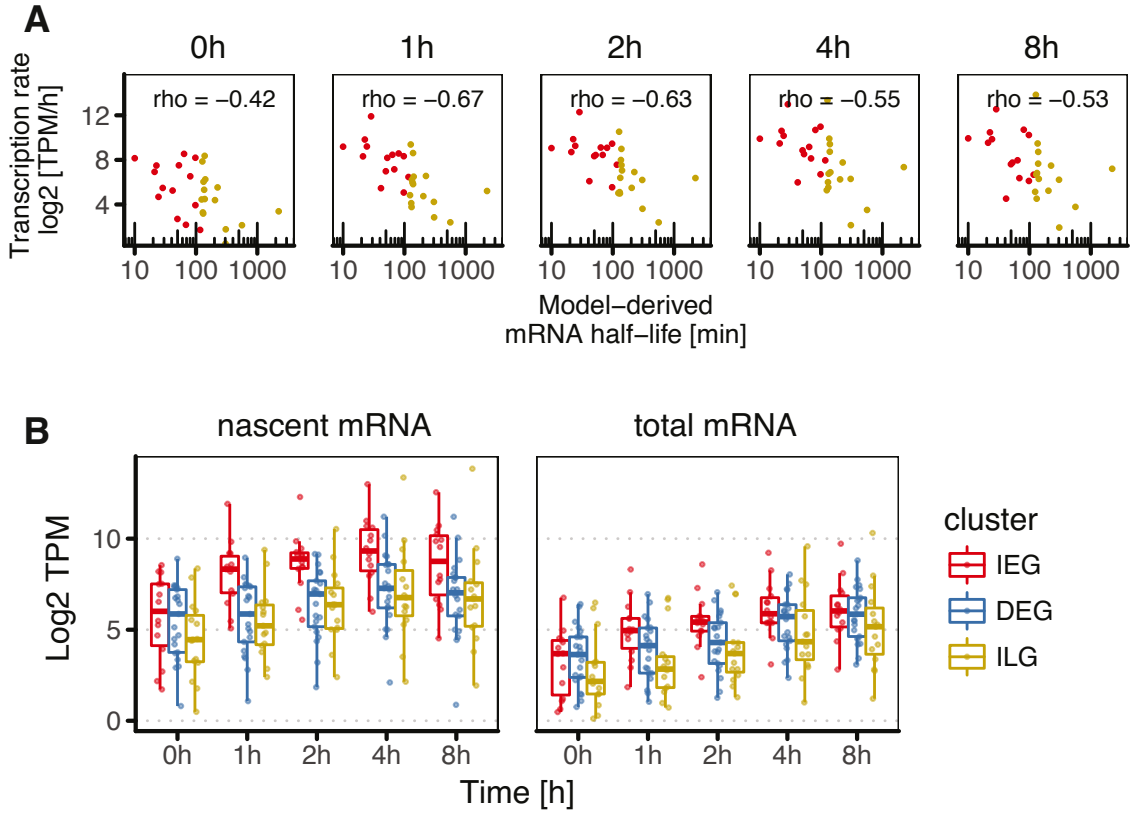


Figure 3.10 Anti-correlation of transcription rate and mRNA half-life and comparison of nascent and total mRNA levels in 4OHT-stimulated HEK293ΔRAF1:ER.
A: Comparison of absolute $\log_2$ transcription rate [TPM/h = transcripts per million per hour] after different periods of 4OHT treatment and model-derived mRNA half-life in HEK293ΔRAF1:ER. Anti-correlation indicates that short mRNA half-lives in IEGs are compensated with high transcription rates.
B: Comparison of nascent and total mRNA levels after different periods of 4OHT treatment in HEK293ΔRAF1:ER. IEGs have higher nascent mRNA levels after stimulation than ILGs and DEGs, but end up at similar total mRNA levels after prolonged activation.

IEGs showed the steepest changes in transcription rate, with a four-fold increase in transcription only one hour post induction. As mRNA levels are determined by the ratio of mRNA production and decay rate, this suggested that IEGs compensate their short half-lives with high transcription rates and may therefore reach similar steady-state levels as long-lived ILGs.

To examine the relation between transcription rates and mRNA half-life, a correlation analysis was performed. Indeed, measured transcription rates (TPM/h = transcripts per million per hour) showed moderate anti-correlation with model-derived half-life estimates for IEGs and ILGs (fig. 3.10A). Anti-correlation was increased in induced samples compared

to untreated samples, suggesting that compensatory effects are pronounced after activation of ERK signalling. Comparison of nascent mRNA and total mRNA levels further confirmed that IEGs can reach high total mRNA levels after both short and prolonged ERK activity (total RNA in fig. 3.10B), by compensating their short half-lives with very high transcription rates (nascent RNA in fig. 3.10B). Interestingly, nascent and total mRNA levels further revealed that ILGs have low basal transcription rates and show low basal mRNA expression levels before stimulation, but reach similar levels as IEGs and DEGs after prolonged activation (fig. 3.10B).

In summary, this chapter identified immediate-late genes (ILGs) as a novel temporal cluster responding to ERK signalling with distinct dynamic properties. Two dynamic features suggested by mathematical modelling were experimentally validated. First, longevity of ILG mRNAs was confirmed by two independent experimental approaches, namely transcriptional shutdown and metabolic labelling experiments. Secondly, promoter GC content indicated a permissive promoter architecture for ILGs and examination of transcription rates confirmed immediacy of ILG transcription upon ERK signalling. The next chapter will focus on investigating implications of these dynamic properties for the decoding of ERK signal duration and aim to put findings into perspective of cell fate decisions in HEK293ΔRAF1:ER cells and other prominent model systems for the study of ERK signal duration.

4 ERK signal duration decoding by mRNA dynamics

Thinking in metaphors may be a hazardous way of drawing scientific conclusions, but combined with numerical simulations and experimental anchors, it may well be the best way the human mind can usefully grapple with biological complexity.

—Upinder S. Bhalla (2003)

4.1 Introduction

Scientific language is packed with metaphors. In the history of cell biology, two main metaphoric conceptualisations have guided the production and organization of knowledge (Reynolds 2018, p. 4). They were adopted from sociology and engineering and rendered microscopic observations and molecular processes tangible.

When studied in context of an organism or simply in relation to other cells in their environment, cells have been described as *social agents* that *communicate* with one another, show certain *behaviours*, form *societies* and are capable of *division of labour*. Interestingly, Rudolf Virchow described the animal body as a free *state of individuals* or a *democratic cell state*, whereas his student Ernst Haeckel discriminated between primitive *cell republics*, referring to plants, and more advanced and tightly controlled *cell monarchies*, referring to animals (Reynolds 2018, pp. 25-30), clearly reflecting their different political views. Nonetheless, the notion of the body as a society of cells and of cells as individual citizens therein helped scientists not only to better communicate their findings with the public, but evoked questions on cellular hierarchies, interactions and individual misbehaviours that might account for pathologies of the encompassing organism (Virchow 1859).

At the same time, with the cell being identified as the fundamental building block of life, a reductionist approach to biological problems was stimulated (Mazzarello 1999). In turn, the scientific language of cell biology was complemented with more mechanistic metaphors. When studied individually, cells were described as chemical *laboratories*, *factories* or complex *machines* which consist of certain *components* or *modules*. In other words, technomorphic/mechanistic metaphors were favoured over anthropomorphic/agential ones (Reynolds

2018, p. 115). At first, the notion of the cell as a machine suggested that it could be broken down into functional parts which in turn could be engineered in a way to improve or impair their efficiency or productivity, or that cells could be entirely reconstructed from scratch, an idea that later gave rise to the field of synthetic biology and which has been referred to as the "engineering ideal" in biology (Pauly 1987).

At most with the emergence of signal transduction research it became evident that holistic cell theories would be reliant on the use of terminology adopted from both sociology and engineering. In fact, the notion of signal transduction was popularized when concepts from cybernetics and information theory were introduced to endocrinology by pioneers like Oscar Hechter or Martin Rodbell who were both influenced by cybernetics originator Norbert Wiener (Reynolds 2018, p. 119). Prior to this, endocrinologists commonly used agential metaphors, for example to describe that chemical messengers (hormones) would be distributed over the blood stream and govern the behaviour of cells perceiving that message. The identification of entire signalling networks, however, led researchers to complement their illustrations with terms borrowed from electronic engineering and information theory. It was stated that cells could send, transmit, receive, and, along that way, encode and decode molecular messages. This conceptualisation of cell signalling as a process of communication and the metaphoric terminology established for it were certainly influenced by Claude E. Shannon's mathematical theory of communication (Shannon 1948). Of course, Shannon's theory did not only provide a terminological framework, but, first and foremost, was adopted to study cell biology in general and cell signalling in particular (reviewed in Adami 2004; Waltermann and Klipp 2011; Rhee, Cheong, and Levchenko 2012).

Eventually, any attempt to describe the processes and events of cellular signalling falls back on the mingling of agential and mechanistic metaphors. And this rule does not stop at the ERK signalling network. Cells sense activating growth factors like EGF, FGF or NGF with their receptors, and translate these incoming messages into appropriate cellular responses with help of molecular sensors. The ERK signalling network controls many different cellular responses such as proliferation, differentiation and cell death (Oda et al. 2005). As discussed in the general introduction, it has been shown that the different fates are encoded by signal duration of the terminal kinase in the pathway, ERK (Marshall 1995). Cells commonly interpret transient ERK signalling as a proliferative signal. In contrast, when exposed to sustained ERK signalling, cells can differentiate or undergo cell death in a cell line-dependent manner.

Whereas many molecular mechanisms of ERK signal duration encoding have been deciphered (cf. sec. 1.3), mechanisms responsible for signal duration decoding remain to be thoroughly elucidated (Blüthgen and Legewie 2008; Purvis and Lahav 2013). Hence, the

central question of my thesis introduced on the first page remains open: How are the molecular messages a cell perceives and encodes in its signalling dynamics eventually translated to cellular responses, or more precisely, to cellular phenotypes? And in particular, how can ERK signal duration alter the composition of responding gene sets, which in turn determine cellular fate? In light of the discussed role of metaphors in cell biology in general and in signal transduction research in particular, it is worth revisiting these questions and to ask how they could be rephrased when all metaphors were to be left behind in this case. Clearly, a complete omission of metaphors would be an endeavour difficult to achieve and inevitably result in a formal and probably incomprehensible description of chemical reactions. Nevertheless, it would potentially lead to questioning the underlying physico-chemical principles of the observed phenomena and one might ultimately arrive at the question:

What molecular entity can function to decode signalling dynamics and which physico-chemical property facilitates this function?

In this chapter, I will try to shed light on this question and investigate whether certain mRNAs can function as molecular decoders and examine the role of mRNA half-lives as an immanent property of mRNAs to govern the timing of gene expression modules and to facilitate ERK signal duration decoding. To better illustrate my findings, different ERK signalling dynamics and the resulting gene expression dynamics will be revisited, but this time focusing on differences in response amplitude across the responding gene expression modules that were identified in the previous chapter. Next, the relation between signal duration and response dynamics will be extensively discussed for a set of prominent response genes, namely for immediate-early responders *EGR1* and *FOS* as well as for immediate-late responders *CLU* and *FOSL1*. Afterwards, a global assessment of signal duration decoding capacity with respect to mRNA half-life is performed and the overall importance of mRNA half-life for gene expression timing upon different ERK signal durations is examined. Lastly, the relation between ERK signal duration and resulting mRNA dynamics is validated in other model systems, namely in PC12 and MCF7 cell lines, and functional implications in HEK293ΔRAF1:ER cells are addressed with help of gene term enrichment analyses, flow cytometry experiments and live cell imaging–based phenotypic profiling.

As a final introductory remark to this chapter, it is important to distinguish the concept of signal decoding from a mere relay of signals. More precisely, one must note that the information flow from signalling input dynamics to gene expression output dynamics can be of two kinds. On the one hand, temporal dynamics of signalling elements can be relayed to mRNA expression dynamics, meaning that, for instance, changes in signalling amplitude precisely reflect changes in mRNA expression amplitude or changes in signalling duration precisely

reflect changes in mRNA expression duration. In this scenario of a relayed signal, the task of signal decoding is only postponed to the next level, from mRNA dynamics to protein dynamics. On the other hand, temporal dynamics can be truly decoded, for instance, when changes in signalling duration are translated into mRNA expression amplitude. Although a sharp distinction between signal decoding and signal relaying once more violates the notion of the aforementioned continuous nature of gene expression, it serves as a pragmatic simplification to identify different prototypic signal translation strategies that genes can be assigned to.

4.2 ILGs translate ERK signal duration into response amplitude

Having quantitatively characterised and validated the parametric properties of IEGs, DEGs and ILGs, a next logical step was to investigate mechanistic and functional implications of these properties. First, response amplitudes upon constant ERK signalling elicited by 4OHT were compared with response amplitudes upon two-hour pulse (4OHT treatment followed by U0126 treatment) and transient signalling (EGF treatment) in transcriptome time-course data (cf. fig. 2.3A for sampling). Median induction curves of the different proposed PRG clusters nicely reflected their different kinetic properties (fig. 4.1A). Again, the rapid induction of IEGs and the subsequent response of DEGs and ILGs was apparent. However, overall accumulation of DEGs seemed to be delayed first but rapid later, whereas overall accumulation of ILGs seemed to be immediate but steadily slow. Remarkably, when considering shortened ERK signal durations, some genes were still rising in expression hours after signalling inputs lapsed (fig. 4.1A). This behaviour was predominantly observed among DEGs and more pronounced upon EGF-mediated ERK signalling than upon two-hour pulse signalling. In contrast, expression of IEGs and ILGs nearly instantly declined once signalling inputs lapsed. This observation was very much in accordance with *in silico* analyses presented earlier (fig. 2.6).

When systematically comparing the relation of signal duration and response dynamics across genes, it was found that IEGs generally relayed signal duration to response duration. Response amplitude however was only partially affected for a fraction of IEGs. For all IEGs a two-hour pulse was sufficient to exceed 50% of response amplitude (fig. 4.1B). Notably, even transient activation with EGF was sufficient to induce half of IEGs (11/21) to this extend. In contrast, response amplitude of ILGs was strongly linked to ERK signal duration. Upon sustained signalling, the majority of ILGs reached response amplitudes between 80% and 100%, similar to IEGs. However, upon two-hour pulse signalling, the majority of ILGs

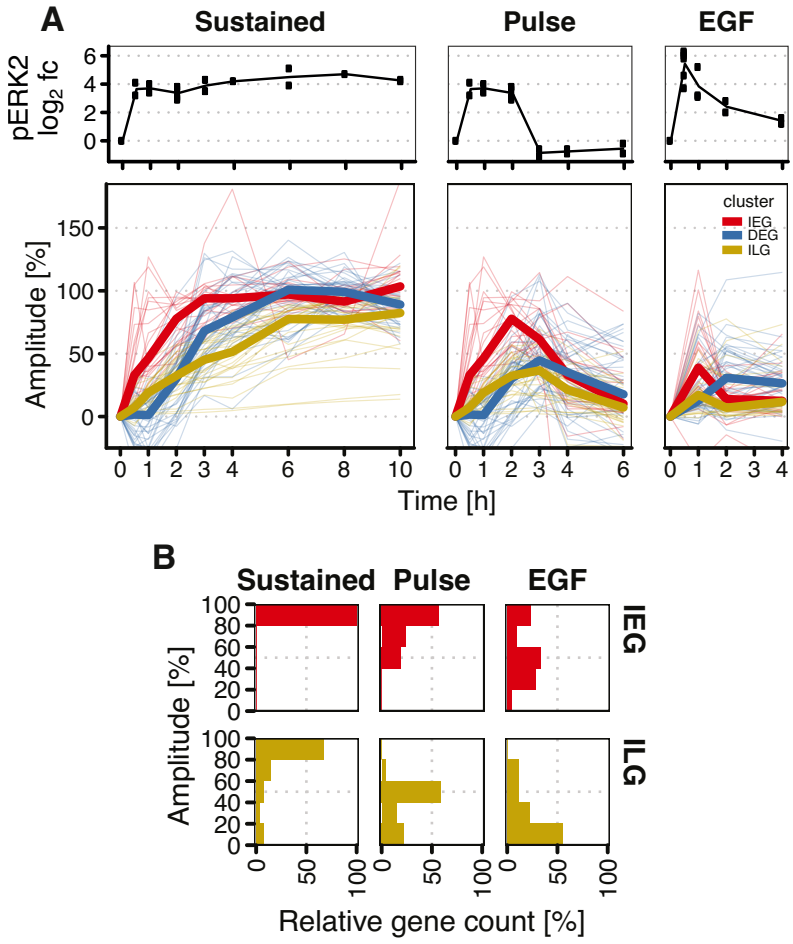


Figure 4.1 Immediate-late genes (ILGs) translate signal duration into response amplitude.
A: Upper panel: pERK2 $\log_2$ fold changes upon different input scenarios (sustained: 4OHT treatment; two-hour pulse: 4OHT plus subsequent U0126 treatment; transient: EGF treatment). Lower panel: Response amplitude across temporal clusters and signal durations as percentage of maximum amplitude of model fits. Bold lines show median cluster amplitude at each time point. Light lines show individual gene trajectories.
B: Binned response amplitudes. Fraction of IEGs and ILGs reaching certain amplitude levels. Amplitude levels are binned in 20% intervals. The largest fraction of ILGs moves from the highest amplitude interval (80%-100%) to the lowest amplitude interval (0%-20%) with respect to signal duration, whereas the majority of IEGs shows response amplitudes greater than 50% across all tested signalling scenarios.

reached response amplitudes only between 40% and 60% and upon transient EGF-mediated signalling the majority of ILGs did not exceed 20% of response amplitude. Thus, ILGs were clearly capable of distinguishing sustained and short signalling by translating signal duration into response amplitude. *In silico* analyses (fig. 2.6) suggested that the capability to decode

signal duration is enabled by longer mRNA half-lives of ILGs compared to short-lived mRNAs of IEGs. In accordance with this, it was found that a fraction of long-lived DEGs with model-derived half-lives greater 120 minutes was also capable of decoding ERK signal duration in a similar manner (*ARL5B, BHLHE40, C2orf42, CDKN1A, GADD45A, GPR50, HOMER1, KDM6B, KRT8, PPP1R15A, SERPINB9, TFPI2, TNFRSF12A, TXNL4B*).

The systematic assessment of the three different signalling input scenarios was complemented by obtaining mRNA and protein measurements on a gene-by-gene basis for a defined set of response genes with help of qRT-PCR and western blotting. Using HEK293ΔRAF1:ER cells, ERK signalling was activated in a series of time course experiments covering five different signal durations. Probed signal durations ranged between thirty minutes and eight hours, and three independent replicates were considered for each time point (cf. fig. 2.3C for sampling). The complete qRT-PCR expression panel consisted of seventeen highly regulated IEGs, ILGs and long-lived DEGs (fig. 4.3A). However, for illustrative purposes, the following discussion focuses on a subset of genes for which corresponding western blots were obtained (fig. 4.2).

Probing of different signal durations confirmed that IEGs like *EGR1* and *FOS* relay signal duration to mRNA response duration, i.e. longer pulses led to a prolonged expression with the same amplitude (fig. 4.2A). ILGs like *CLU* and *FOSL1* on the contrary were found to truly decode signal duration by translating it into response amplitude, i.e. maximum expression levels increased with longer signal duration. Assuringly, the principle was not only apparent when looking at individual genes, but also in the average response amplitude for all qRT-PCR-validated genes (fig. 4.2B). For tested IEGs, the median response amplitude increased only slightly from 0.5 to 1 hour and remained at 100% for longer durations. In contrast, the median response amplitude of tested ILGs increased steadily with higher signal duration. Taken together, the data confirmed that IEGs relay signal duration, whereas ILGs decode signal duration into response amplitude.

Western blot experiments (cf. sec. A.8) allowed assessment of signal duration decoding at the protein level of the discussed IEGs and ILGs (fig. 4.2C and fig. 4.3B). For *EGR1*, it was found that signal duration is not only relayed to mRNA levels, but also to protein levels. Prolonged ERK activation with 4OHT and transient ERK activation with EGF resulted in prolonged and transient changes in EGR1 protein levels respectively, whereas maximum amplitudes were in a similar range. Hence, it was concluded that EGR1 protein is not able to decode ERK signal duration. FOS protein levels, in contrast, suggested strong ERK signal duration decoding. In accordance with the FOS protein sensor model (Murphy et al. 2002), FOS protein accumulated with respect to ERK signal duration. Therefore, it could be concluded that IEGs can potentially decode signal duration on the protein level if the

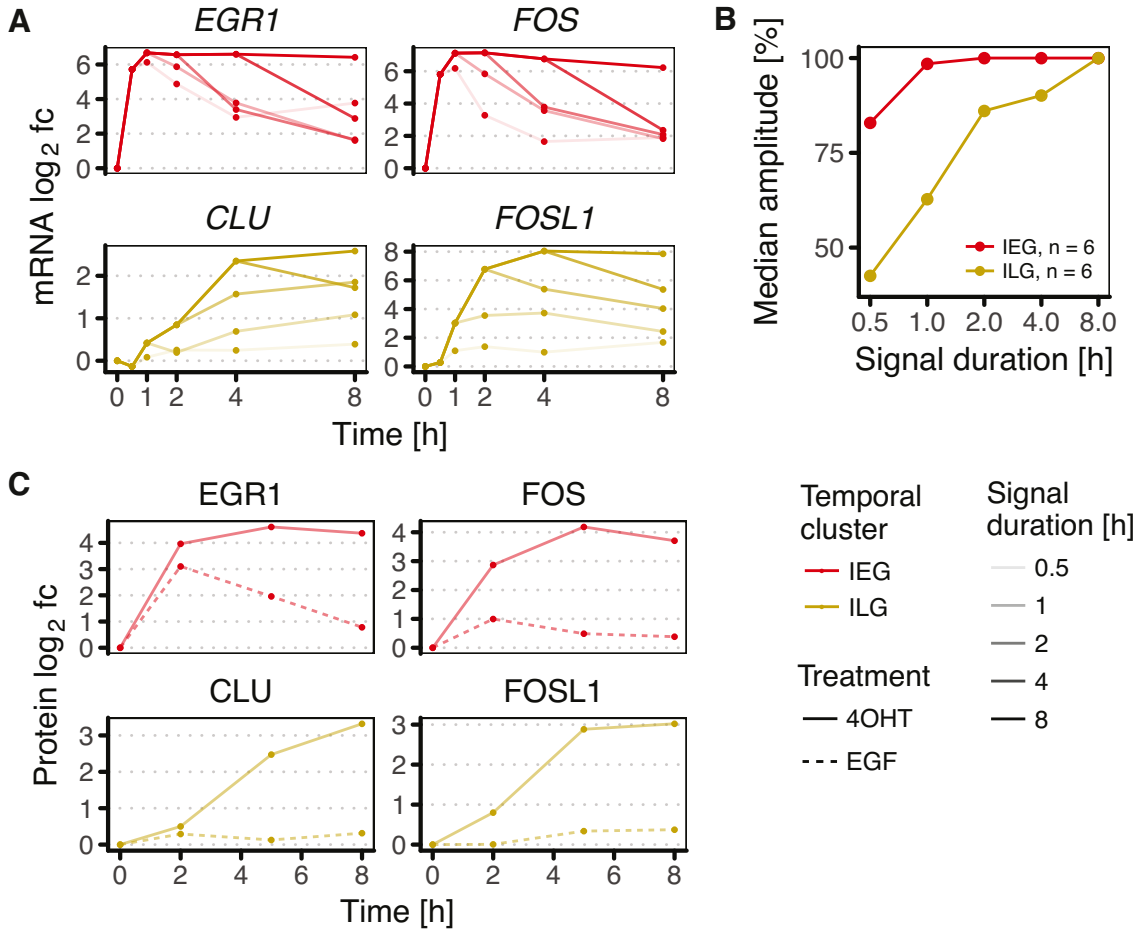


Figure 4.2 Validation of signal duration decoding in a set of example genes. **A**: RT-qPCR validation to test different ERK signal durations. HEK293ΔRAF1:ER cells were treated with 4OHT and U0126 for different periods of time to generate signal duration scenarios of 0.5 to 8 hours (cf. fig. 2.3C). mRNAs of IEGs *EGR1* and *FOS* relay signal duration to response duration, whereas ILGs *CLU* and *FOSL1* decode signal duration to response amplitude. Colours indicate temporal cluster and alpha levels indicate signal duration. RT-qPCR data for all 17 validated mRNAs is shown in fig. 4.3A. **B**: Relation between signal duration and response amplitude for IEGs and ILGs derived from RT-qPCR validation data (duration-to-amplitude profile). Median amplitude is based on all six RT-qPCR-validated ILGs and all six RT-qPCR-validated IEGs shown in fig. 4.3A. **C**: Quantification of western blots to present protein $\log_2$ fold changes of sample genes upon sustained ERK signalling (4OHT-induced) and transient ERK signalling (EGF-induced) in HEK293ΔRAF1:ER cells. Western blot images are shown in fig. 4.3B. Linetype indicates treatment.

protein itself is stable or stabilised (fig. 4.2C). Interestingly, both tested ILGs *CLU* and *FOSL1* were found to not only accumulate at the level of mRNA (with respect to ERK signal duration), but also at the level of their respective proteins (Clusterin and Fra-1). Both

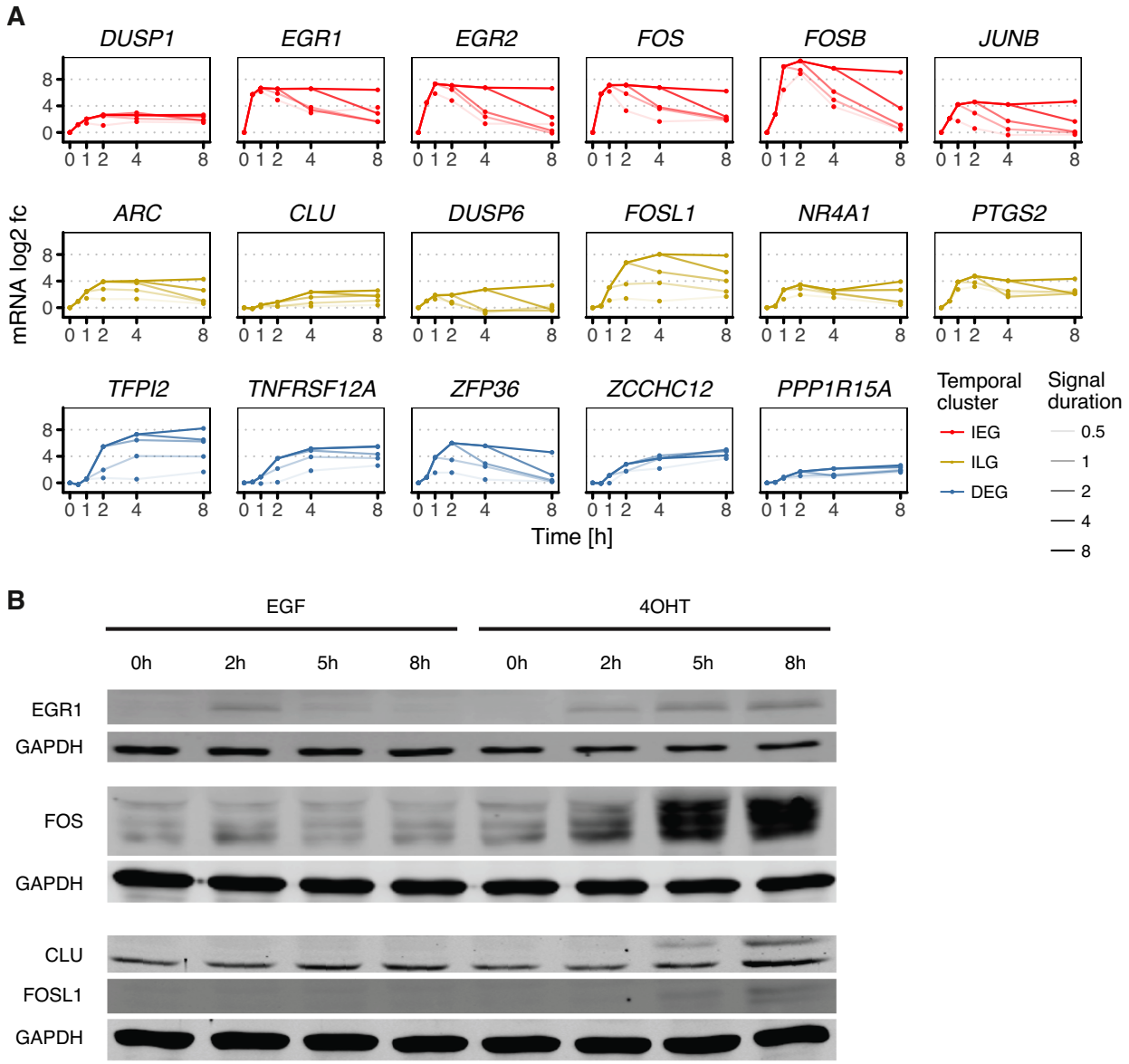


Figure 4.3 Signal duration effects on mRNA and protein level in HEK293ΔRAF1:ER cells.

A: mRNA $\log_2$ fold changes in RT-qPCR time course data for five different ERK signal durations in HEK293ΔRAF1:ER cells (cf. fig. 2.3C for treatment scheme). Colours indicate temporal cluster and alpha level indicates signal duration. Employed primers are listed in sec. A.8.

B: Representative western blot images for protein fold changes upon different periods of EGF and 4OHT treatment in HEK293ΔRAF1:ER cells. CLU and FOSL1 proteins (Clusterin and Fra-1) were measured on the same membrane, hence GAPDH loading control in last row corresponds to both blots. Western blot quantifications are shown in fig. 4.2C.

proteins showed strong differences in response amplitude for the different input stimuli, and were therefore identified to be capable of ERK signal duration decoding. Notably, Clusterin has been described to be a short-lived, poly-ubiquitinated protein (Rizzi et al. 2009). Hence, its identified ERK signal duration decoding capability could not be attributed to a long protein half-life, but potentially results from *CLU* mRNA dynamics being relayed to Clusterin protein dynamics. On the contrary, considering previous reports of ERK-mediated post-translational stabilisation of Fra-1 (Murphy, MacKeigan, and Blenis 2004), the analysis suggested that duration decoding capability of Fra-1 protein potentially results from a mixture of relayed *FOSL1* mRNA dynamics and Fra-1 protein stabilisation.

In conclusion, gene-specific comparisons of mRNA and protein responses suggested four different scenarios of how signal duration can translate to gene product dynamics:

1. **mRNA/Protein relay model**: Signal duration is relayed to the mRNA and protein level if both mRNA and protein are short-lived (EGR1-like signal translation).
2. **Protein decoder model**: Signal duration is relayed to the mRNA level but decoded at the protein level if mRNA is short-lived but protein is long-lived or stabilised (FOS-like signal translation).
3. **mRNA decoder model**: Signal duration is decoded at the mRNA level and decoding capability is relayed to the protein level if mRNA is long-lived but protein is short-lived (CLU-like signal translation).
4. **mRNA/Protein decoder model**: Signal duration is decoded at the mRNA level and decoding capability is complemented at the protein level if both mRNA and protein are long-lived or stabilised (FOSL1-like signal translation).

4.3 mRNA half-lives govern gene expression timing and affect signal duration decoding capacity

So far, different gene expression modules and particular genes of interest were compared and tested for their capability to decode ERK signal duration in a qualitative manner. Once again, it is important to stress that such qualitative assignments can be very helpful to provide a comprehensible description of observations and ease the communication of key findings. At the same time, however, it necessarily implies a discretisation of the aforementioned continuous nature of gene expression. Accordingly, this section now aims to complement the qualitative descriptions of the previous section and proposes a quantitative measure for signal duration decoding, termed signal duration decoding capacity (SDDC).[1]

[1] As opposed to the term of signal duration decoding capability used in the previous section.

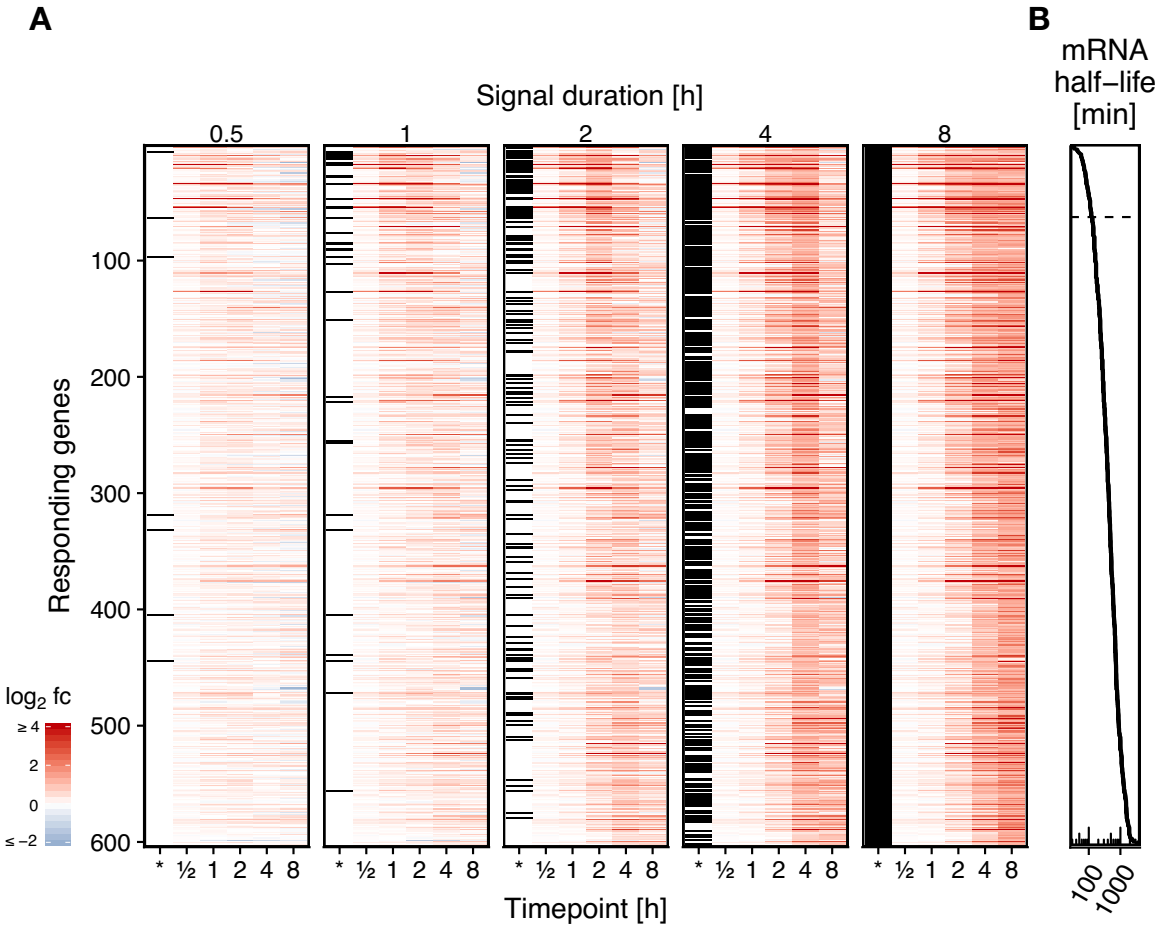


Figure 4.4 mRNA half-lives govern gene expression timing.
A: A colour-coded heatmap of $\log_2$ fold changes shows changes in gene expression of 604 significantly upregulated genes across five different ERK signal durations profiled in HEK293ΔRAF1:ER cells. RNA-sequencing was used to measure gene expression and signal duration was controlled by timed 4OHT and U0126 treatments (cf. fig. 2.3C for sampling). Genes were ranked by median mRNA half-life estimates shown in **B**. Black bars in the first column (*) of each subplot indicate that the respective gene responded with a relative response amplitude of at least 80% upon the given signal duration (corresponding to the highest amplitude bin in fig. 4.1B).
B: mRNA half-life estimates of all 604 induced genes shown in **A**. Inference of estimates is described in section 3.4. Dashed line separates genes with half-lives either smaller than 120 minutes (above line) or greater than 120 minutes (below line).

All analyses presented in this section were based on RNA-sequencing data obtained from the same biological samples as the previously presented RT-qPCR data (cf. fig. 2.3C for sampling).

Before SDDC values were determined, an analysis of differential gene expression was performed. In total, 604 genes were identified to be significantly upregulated upon sustained

4OHT treatment. A comparison of the newly obtained RNA-sequencing data and the previously obtained microarray gene expression time course data showed that 140 out of 189 (74%) initially identified primary and secondary response genes could be verified to be induced upon 4OHT treatment. Here, the 464 additionally identified response genes might be attributed to a higher sensitivity of RNA-sequencing–based profiling compared to microarray-based profiling and the stringent filtering of the obtained microarray data (cf. sec. A.3). To provide a comprehensive presentation of the newly acquired transcriptome data, $\log_2$ fold changes for all 604 induced genes across the five different ERK signal durations were presented in a heatmap (fig. 4.4). Genes were ranked by their estimated mRNA half-life to demonstrate the general importance of mRNA half-life for gene expression timing (cf. sec. 3.4 for mRNA half-life inference). Clearly, the data showed that short-lived genes respond fast and long-lived genes respond at later time points. Upon a two-hour pulse of ERK signalling, more than half of the genes with short mRNA half-lives smaller than 120 minutes (58%) were already considered maximally induced, as they reached relative response amplitudes greater than 80%. Among long-lived genes, in contrast, only a minority (16%) exceeded this threshold after two hours of ERK signalling. After four hours, nearly all short-lived genes (91%) responded to this extend, whereas only two-thirds of long-lived genes did (64%). After eight hours, all genes consequently reached their respective maximum amplitude. Taken together, this initial examination of the acquired RNA-sequencing data once again underlined that mRNA half-lives are an important feature of mRNA dynamics as they govern the timing of gene expression in response to ERK signalling.

In a next step, the signal duration profiling data was used to systematically determine the signal duration decoding capacity for all responding genes. For this, SDDC was derived from the relation between ERK signal duration and the maximum gene response amplitude and was calculated as one minus the area under the curve in a given duration-to-amplitude profile (cf. sec. A.7). By definition, a linear relation between (relative) signal duration and (relative) response amplitude would yield an ideal signal duration decoding capacity of 0.5. A concave non-linear relation would result in lower values indicating a smaller capacity to decode signal duration with values close to zero indicating a near rectangular relation and almost no capacity to do so.

To illustrate SDDC quantification, values were first calculated for the same IEGs and ILGs discussed in the previous section and inference was visualised with help of duration-to-amplitude profiles (fig. 4.5A). The previous assessment of mRNA response dynamics had suggested that IEGs *EGR1* and *FOS* are not capable of decoding ERK signal duration at the mRNA level (mRNA relay model), whereas ILGs *CLU* and *FOSL1* were found to be capable to do so (mRNA decoder model). And indeed, quantification of duration-to-

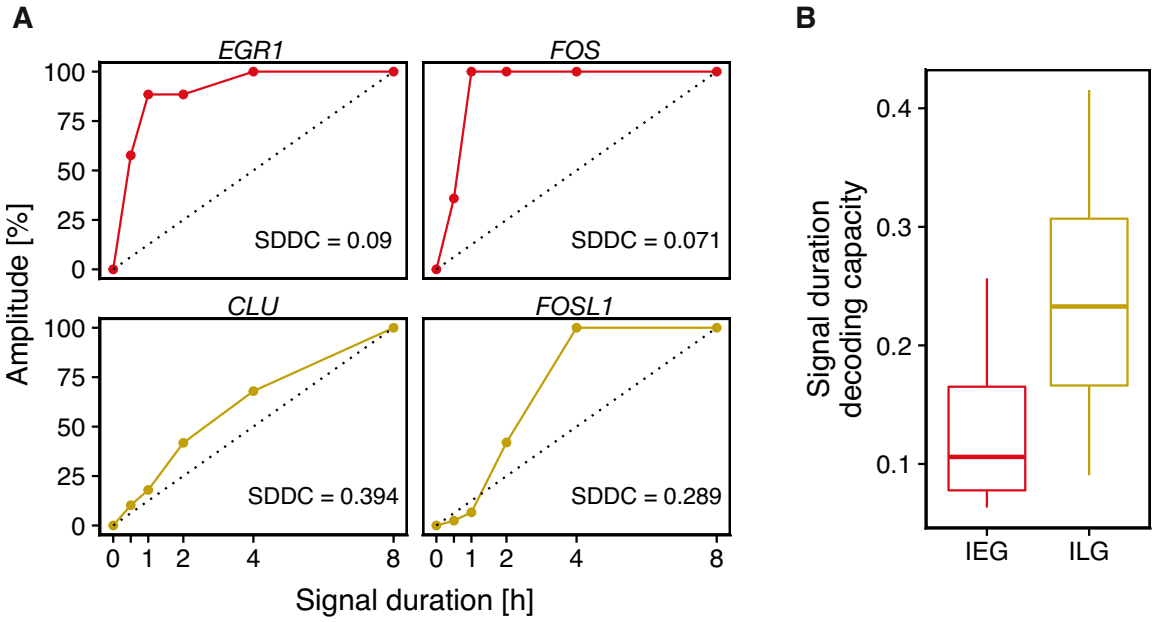


Figure 4.5 Different signal duration decoding capacities (SDDC) in IEGs versus ILGs.
A: Gene-specific signal duration decoding capacities can be inferred by assessing the maximum response amplitude for a given signal duration (depicted as duration-to-amplitude profiles). SDDCs of IEGs *EGR1* and *FOS* are smaller than SDDCs of ILGs *CLU* and *FOSL1*. Dotted lines indicate ideal linear relation between signal duration and response amplitude and correspond to an optimal SDDC of 0.5.
B: Box plot distributions of signal duration decoding capacities show that ILGs have a higher median SDDC than IEGs.

amplitude profiles yielded very small SDDC values for *EGR1* and *FOS* (0.09 and 0.071 respectively), whereas SDDC values for *CLU* and *FOSL1* were notably larger (0.394 and 0.289 respectively). A comparison of SDDC values for all RNA-sequencing–validated IEGs (n = 16) and ILGs (n = 23) confirmed gene-specific observations (fig. 4.5B). Overall, IEGs showed a median SDDC of 0.1, whereas ILGs showed a median SDDC of 0.24.

Higher SDDC values in ILGs compared to smaller SDDC values in IEGs again suggested a potential role for mRNA half-life to affect signal duration decoding capacity. Hence, in a next step, the relation between mRNA half-life and SDDC was assessed in a systematic fashion by subdividing all responding genes into five separate bins based on their mRNA half-life (fig. 4.6). Four-hour intervals were used for initial binning and the first interval was further subdivided into two two-hour intervals. As a result, the first bin consisted of 62 response genes with mRNA half-lives smaller than 120 minutes and thus corresponded to an extended group of IEGs, as IEGs had been defined to have mRNA half-lives smaller

than 120 minutes (cf. sec. 3.2). All other bins consisted of 107 to 169 genes with longer mRNA half-lives and corresponded to groups of late responding genes. However, since PBS dependency was not assessed in this experiment, a discrimination of PRGs and SRGs could not be made and, in turn, a clear assignment of the additionally identified response genes to a particular gene expression module could not be provided.

In agreement with temporal cluster–specific assessments (fig. 4.2B) the set of short-lived genes showed the steepest increase in response amplitude with respect to signal duration (fig. 4.6A). On average, genes in this group exceeded 50% response amplitudes after one hour and were almost maximally induced after four hours (95%), hinting at a weak capacity of signal duration decoding. And indeed, SDDC values were smallest for this group with a median value of 0.16. With an increase of mRNA half-life, duration-to-amplitude profiles became consistently less concave. Notably, median SDDC values generally increased with respect to an increase of mRNA half-life. Across bins, median SDDC increased up to a value of 0.32 for the set of genes with mRNA half-lives greater than eight hours (fig. 4.6B). Lastly, rank correlation analysis confirmed a positive relation between signal duration decoding capacity and mRNA half-life (Spearman's rho = 0.429, fig. 4.6C).

Based on these observations, it could be concluded that mRNA half-lives clearly affect duration-to-amplitude decoding capacity. However, it should be noted that a generally high and steadily increasing variation of SDDC values was observed across gene sets. Interquartile ranges increased with respect to mRNA half-life (fig. 4.6B) and the generally large and steadily increasing variation of SDDC values was also apparent when bins were omitted to present the data in a continuous scatter plot (fig. 4.6C). A large fraction of the observed variance could potentially be attributed to technical and biological noise in the obtained gene expression data, as well as to uncertainty of the provided mRNA half-life estimates. In addition, SDDC values may not be meaningful for certain genes which are incompatible with the underlying assumptions of the proposed measure. More precisely, one should note that the idealised relation between signal duration and median amplitude presented here assumes an immediate one-step kinetic of gene induction. As discussed earlier (cf. sec. 1.1 and 3.4), many response genes like DEGs and SRGs are regulated by positive feed-forward motifs and follow different types of multi-step induction kinetics which may result in sigmoidal or even convex duration-to-amplitude profiles. The latter case would still result in meaningful SDDC values between 0.5 and 1, indicating a strong decoding capacity with values close to 0.5 and a weak decoding capacity with values close to 1. In case of sigmoidal profiles, however, SDDC values could be very much misleading. For example, a hypothetical switch-like response could occur at an intermediate signal duration and yield a SDDC value close to 0.5, but would otherwise not be considered to be capable of signal duration decoding, since the continuous

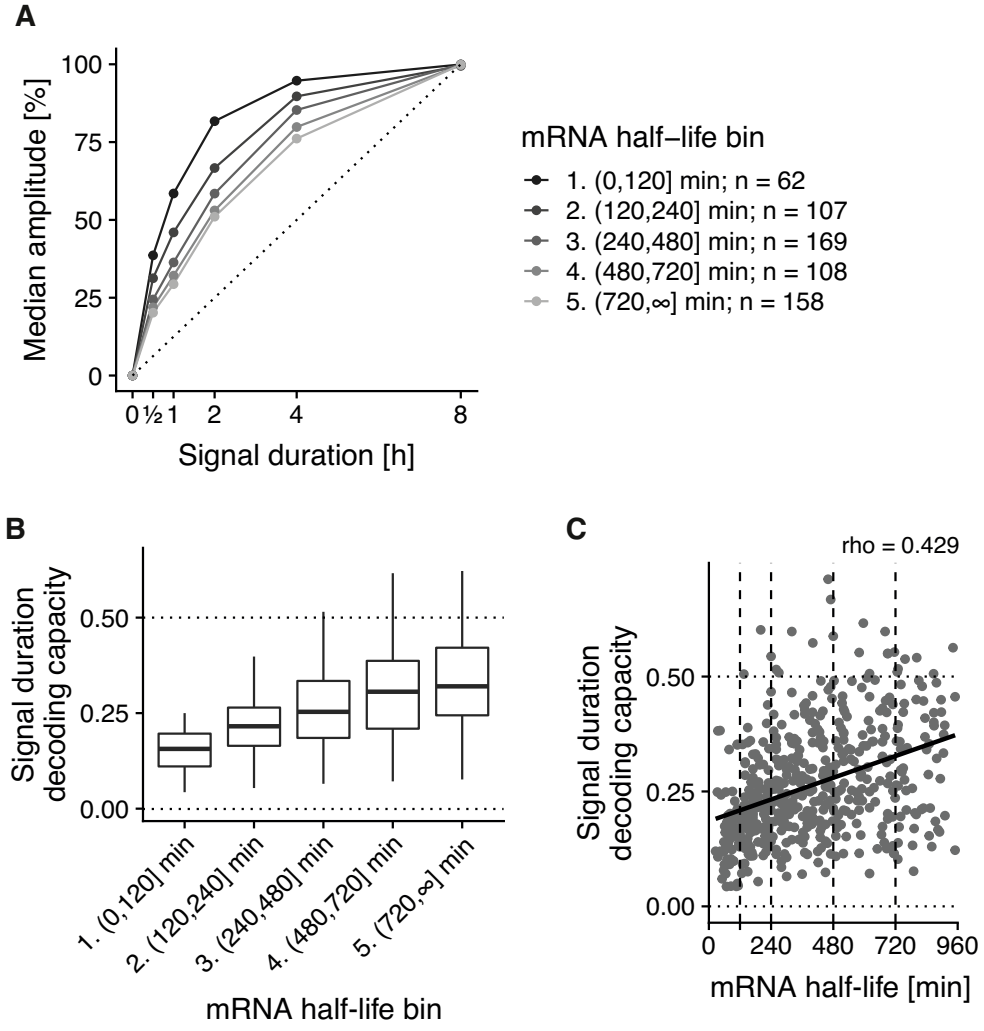


Figure 4.6 mRNA half-lives affect signal duration decoding capacity (SDDC).
A: Duration-to-amplitude profiles for different sets of response genes grouped by mRNA half-life. Grey scale indicates mRNA half-life bin and solid points indicate sampled signal durations. Dotted line denotes ideal linear relation between (relative) median amplitude and (relative) signal duration. n: number of genes per bin.
B: Box plots of signal duration decoding capacity distributions for different mRNA half-life bins show that SDDC increases with respect to mRNA half-life. Dotted lines indicate theoretical minimum and maximum for SDDC.
C: Scatter plot of SDDC for all responding genes with respect to their mRNA half-life illustrates positive correlation of the two parameters (Spearman's rho = 0.429). Dashed vertical lines indicate mRNA half-life bins used in **A** and **B**, dotted horizontal lines indicate theoretical minimum and maximum for SDDC, solid line denotes linear fit of the data.

signalling input is compressed and reduced to a binary all-or-none response output in this scenario. Both ways, duration-to-amplitude profiles for genes following multi-step induction kinetics would be affected by a mixture of mRNA half-life and the number of steps required for their induction. Hence, a more stringent selection of response genes including only those following an immediate one-step induction and showing only low levels of expression noise would potentially be more suitable to examine the effect of mRNA half-life on signal duration decoding capacity.

To conclude this section, a systematic profiling of mRNA dynamics in response to a wide range of different synthetically controlled ERK signal durations confirmed that mRNA half-life is an important feature of mRNA dynamics which governs gene expression timing and strongly affects signal duration decoding capacity of response genes. However, a high variation in SDDC values with respect to mRNA half-life suggested that additional features like the number of steps required for gene induction could impair the predictive power of mRNA half-life to forecast the capacity of a certain gene to decode signal duration.

4.4 ERK signal duration decoding and gene expression timing is conserved in PC12 and MCF7 cells

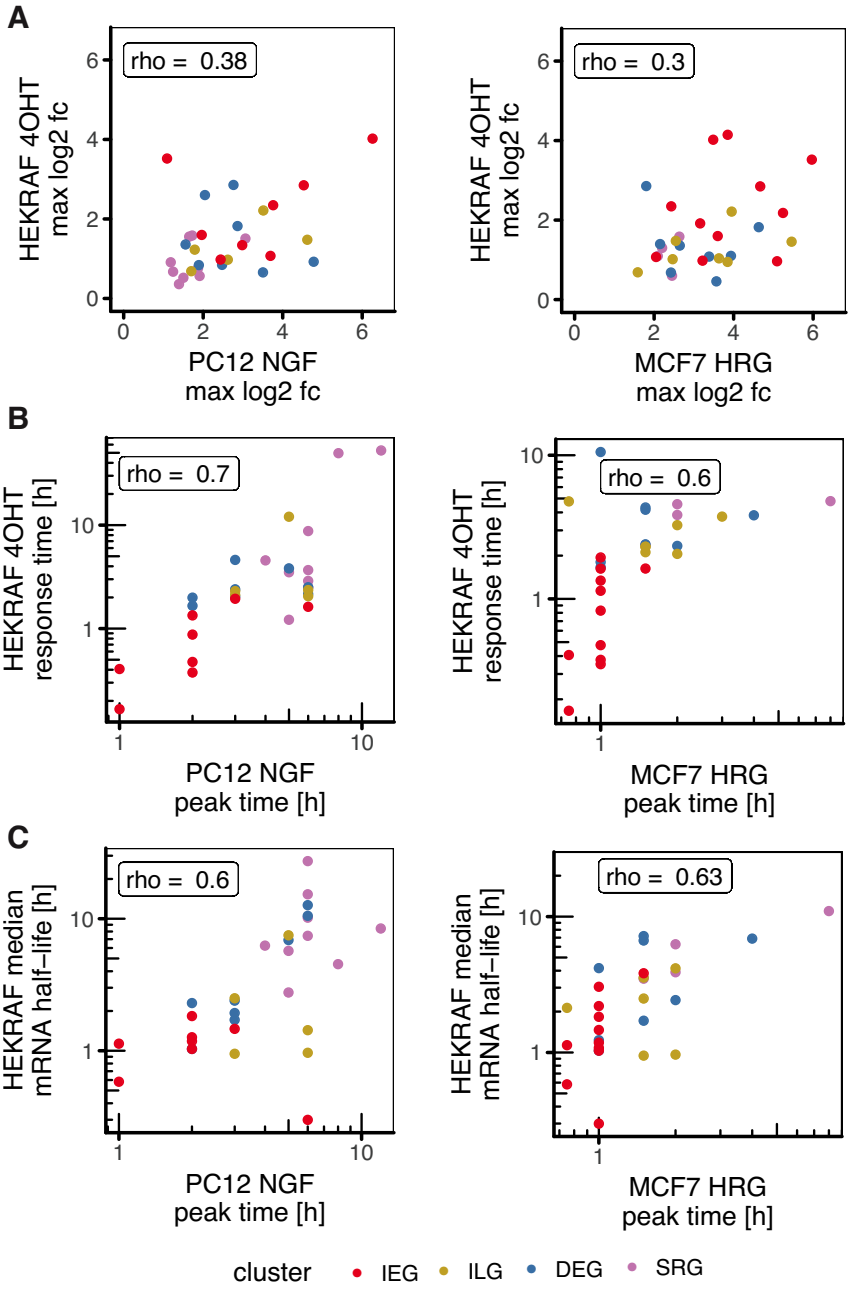
As elaborated in the general introduction of my thesis, a relation between signal duration and cell fate decisions was first described for rat PC12 cells (Traverse et al. 1992). In these cells, transient ERK activity elicited by epidermal growth factor (EGF) resulted in proliferation, whereas sustained ERK activity elicited by neuronal growth factor (NGF) resulted in cellular differentiation. The relation between ERK signal duration and cell fate decisions was later confirmed for many other systems (cf. sec. 1.3). For example, it was found that MCF7 cells undergo proliferation or commit to apoptosis when exposed to transient or sustained ERK signalling elicited by EGF or heregulin (HRG) respectively. Today, both PC12 and MCF7 cells commonly serve as paradigm model systems to study ERK signal duration. Using these and other model systems it was established that, mechanistically, ERK signal duration is dictated by a range of negative feedback mechanisms acting on different nodes of the ERK signalling network. In EGF-mediated ERK signalling, ligand-induced receptor endocytosis leads to EGFR degradation and prevents sustained activation of ERK (Avraham and Yarden 2011). Furthermore, negative feedbacks from ERK to EGFR, SOS, Ras and RAF can account for transient dynamics in EGF-treated cells (Lake, Corrêa, and Müller 2016). In contrast to EGF, other growth factors such as NGF and HGF inhibit endosomal fusion, thereby prolonging the life-time of EGFR and promoting receptor recycling to the plasma membrane (Villaseñor et al. 2015). Also, negative feedbacks from ERK can potentially be evaded by

signalling through RAP1 (Sasagawa et al. 2005) or by promoting sustained signalling via PKC-mediated positive feedback control of RAS (Bhalla, Ram, and Iyengar 2002; Santos, Verveer, and Bastiaens 2007).

So far, nearly all analyses presented in my thesis were based on measurements obtained from a highly controllable synthetic cell culture system (HEK293ΔRAF1:ER). Although different growth factors (EGF and FGF) were used to elicit different physiological ERK signalling dynamics in this system, most of the profiled signalling dynamics were synthetically induced with help of a controllable fusion protein to generate a defined set of different ERK signal durations. In consequence, the described feedback mechanisms which account for signal duration differences in physiological ERK signalling were rendered moot in this system. Thus, it is important to test whether the conclusions presented so far hold true for more physiological model systems like NGF/EGF–treated PC12 cells or HRG/EGF–treated MCF7 cells. Along these lines, this section systematically examines the role of mRNA half-life for gene expression timing and for decoding ERK signal duration in these two model systems.

The gene regulatory response to different ERK signal durations in PC12 and MCF7 cells was assessed using two publicly available data sets (Offermann et al. 2016; Saeki et al. 2009, cf. sec. A.10). The data sets consist of gene expression time course measurements obtained from PC12 cells treated with NGF or EGF and MCF7 cells treated with HRG or EGF. First, a list of differentially expressed genes was determined for all cell lines and treatments. Next, each list was filtered to only include genes that had been identified to respond to ERK signalling in HEK293ΔRAF1:ER cells. For PC12 cells, rat homologs were considered for gene matching. Notably, 30 out of 189 previously identified response genes were also

Figure 4.7 ***(facing page)*** **Conservation of mRNA response dynamics in rat PC12 and human MCF7 cells.**
A: Spearman correlation of maximum mRNA $\log_2$ fold changes in 4OHT-treated HEK293ΔRAF1:ER cells and maximum mRNA $\log_2$ fold changes of corresponding homologues in NGF-treated PC12 cells (left panel) and HRG-treated MCF7 cells (right panel).
B: Spearman correlation of mRNA response times in HEK293ΔRAF1:ER cells and peak expression time points in PC12 and MCF7 cells.
C: Spearman correlation of median mRNA half-lives in HEK293ΔRAF1:ER cells and peak expression time points in PC12 and MCF7 cells.
D: Conservation of signal duration decoding to response amplitude in PC12 and MCF7 cells: Amplitude ratios correspond to ratios of maximum fold changes upon NGF and EGF or HRG and EGF induction. In both cell lines, decoding of signal duration to response amplitude was found to be governed by ILGs. Average differences of amplitude ratios between IEGs and ILGs were significant for MCF7 cells (p-value = 0.011; Wilcoxon rank sum test).


A
rho = 0.38
HEKRAF 4OHT max log2 fc
PC12 NGF max log2 fc
rho = 0.3
MCF7 HRG max log2 fc
B
rho = 0.7
HEKRAF 4OHT response time [h]
PC12 NGF peak time [h]
rho = 0.6
MCF7 HRG peak time [h]
C
rho = 0.6
HEKRAF median mRNA half–life [h]
PC12 NGF peak time [h]
rho = 0.63
MCF7 HRG peak time [h]
cluster IEG ILG DEG SRG
D
PC12 NGF / EGF
MCF7 HRG / EGF
Amplitude ratio (log2)
IEG ILG
IEG ILG

responding to sustained ERK signalling in PC12 and MCF7 cells respectively. For both cell lines, the consensus list of response genes included more PRGs than SRGs. For PC12 cells 21 PRGs (8 IEGs, 8 DEGs, 5 ILGs) were accompanied by 9 SRGs, and for MCF7 cells 26 PRGs (11 IEGs, 8 DEGs, 7 ILGs) but only 4 SRGs were part of the consensus list. This observation was in line with the established idea of PRGs being more commonly shared across cell types and treatments and SRGs responding in a rather context-specific manner (Avraham and Yarden 2011). Based on the consensus gene sets, four different comparisons with data from HEK293ΔRAF1:ER cells were made to test the conservation of different features of mRNA response dynamics in general and the identified signal duration decoding principle in particular (fig. 4.7A-D).

In a first comparison, maximum $\log_2$ fold changes were assessed as a prominent feature of mRNA response dynamics (fig. 4.7A). Overall, correlation of NGF- and HRG-induced mRNA $\log_2$ fold changes in PC12 and MCF7 cells with 4OHT-induced mRNA $\log_2$ fold changes in HEK293ΔRAF1:ER cells was only weak (Spearman's rho of 0.38 and 0.3 respectively). Here, one should note that maximum fold changes upon sustained ERK signalling do not provide any information about gene expression timing, but only result from relative changes in transcription rate (assuming a minimal model of gene expression with constant mRNA degradation rate). Hence, the weak correlation of fold changes suggested that ERK signalling–mediated changes of transcription rates strongly vary across cell lines and treatments.

Next, conservation of gene expression timing was directly tested by comparing two different parameters of mRNA dynamics derived from HEK293ΔRAF1:ER cells to peak expression time points in PC12 and MCF7 cells (fig. 4.7B+C). Interestingly, both model-derived response times and median mRNA half-life estimates correlated nicely with peak expression time points in PC12 and MCF7 cells (Spearman's rho between 0.60 and 0.70). Together, both comparisons suggested that the temporal aspects of mRNA dynamics in response to ERK signalling are strongly conserved across model systems and treatment conditions. When comparing this finding to the weak correlations of mRNA fold changes (fig. 4.7A), it could be concluded that the timing of the gene regulatory response to ERK signalling is markedly higher conserved than its magnitude.

Lastly, it was tested whether the identified relay of ERK signalling dynamics by IEGs and the contrasting capability of ILGs to decode signal duration could be validated in PC12 and MCF7 cells. For this, amplitude ratios were calculated for IEGs and ILGs by dividing maximum response amplitudes upon NGF/HRG treatment by maximum response amplitudes upon EGF treatment in both systems. Here, it was found that amplitude ratios in both PC12 and in MCF7 cells were higher for ILGs when compared to IEGs. For MCF7 cells, the

difference in amplitude ratios was statistically significant (p-value = 0.011; Wilcoxon rank sum test). Based on this analysis, it was concluded that ILGs are more capable to govern signal duration decoding than IEGs across all tested models for ERK signal duration.

Altogether, the examination of published gene expression time course data sets obtained from different ERK signal duration model systems and comparisons with data generated from HEK293ΔRAF1:ER cells suggested that both gene expression timing and signal duration decoding by long-lived genes is conserved in all three model systems, whereas mRNA $\log_2$ fold changes were found to correlate only weakly.

4.5 ILGs might serve as a fail-save mechanism to control aberrant ERK signalling

Cell signalling involves a complex cascade of events. An external molecular message is encoded in a cellular signal, decoded by a response element and ultimately results in a cell fate decision. As mentioned, stimulus-to-signal encoding has been studied extensively. Hence, this study focused efforts on investigating the mechanism of signal-to-response decoding. Cell signalling, however, does not stop there. Eventually, the decoded molecular message has to be put into action, i.e. a cell fate decision has to be made. Thus, this final section of the chapter aims to phenotypically examine the response to transient and sustained ERK signalling in HEK293ΔRAF1:ER cells with help of flow cytometry and live cell imaging, and to investigate the functional role of the identified ERK-regulated gene expression modules to control the observed cell fates.

At this point, to further motivate the analyses presented here, it is helpful to more carefully revisit Shannon's mathematical theory of communication (Shannon 1948) as a mechanistic metaphor for cell signalling and to discuss the common and widely used terms of *response gene(s)*, *response element(s)* or the term *gene response* in general. Throughout my thesis, these terms were used to describe particular genes or entire groups of genes which would change their expression upon activation of ERK signalling. If one were to push the adoption of information theory for cell signalling to the very extreme, however, one could argue that this use of terms is slightly inaccurate. For example, let growth-factor identity be the mean of communication by which a secretory cell can send the molecular *message* for its neighbouring cells to divide or differentiate. And let a responding cell's signalling network be its *transmitter/encoder* element. In consequence, genes regulated by the signalling network that are capable of decoding the molecular information, should rather be seen as cellular *receiver/decoder* elements, since the actual *response* to the received message only materialises when a responding cell eventually undergoes cell division or differentiation respectively. Cer-

tainly, the question of signalling-regulated genes being either seen as responding or receiving elements in context of cellular communication cannot be answered conclusively, as it simply depends on what one defines as the message to be interpreted by a cell. Nonetheless, in case of the newly identified gene expression module of immediate-late genes (ILGs), it is worth challenging the established terminology and to consider it as a group of cellular receivers rather than responders. Firstly, it is fruitful to introduce this specification as it better underlines the role of ILGs to decode signal duration and, secondly, it stresses that one must not forget to study whether the decoding process is followed by an actual cellular response. In other words, a "thinking in [mechanistic] metaphors" (Bhalla 2003) here emphasises that the success of cellular encoding and decoding processes can only be measured if an appropriate cellular response can be experimentally detected.

To experimentally confirm the different phenotypes reported for HEK293ΔRAF1:ER cells in response to transient and sustained ERK signalling (Cagnol, Obberghen-Schilling, and Chambard 2006), two different experimental methods were considered. In a first attempt, cells were either only treated with 4OHT to elicit sustained ERK signalling or additionally treated with U0126 two hours after the initial 4OHT treatment to elicit a two-hour pulse of ERK signalling (fig. 4.8A). 48 hours post treatment cells were stained with a fluorophor-linked antibody against cleaved caspase-3 to biochemically mark apoptotic cells. Marked cells were then identified and quantified with help of flow cytometry (cf. sec. A.9). Upon sustained ERK signalling, 19.6% of cells were found to be positive for cleaved caspase-3 and were hence considered apoptotic. In contrast, upon transient ERK signalling only 1.7% of cells were found to be apoptotic and a comparable amount of 1.2% of EtOH-treated control cells were marked positive. These observations confirmed aforementioned reports that only sustained ERK signalling is sufficient to induce apoptosis in HEK293ΔRAF1:ER cells, whereas the subsequent U0126-mediated inactivation of ERK signalling prevents cells from undergoing apoptosis.

In a second validation experiment (fig. 4.8B), cells were exposed to the same treatment regime again, but this time continuously monitored with help of a live cell imaging platform (IncuCyte ZOOM, cf. sec. A.9). Here, caspase-3 and caspase-7 activity was detected with help of a conditionally fluorescent reagent that releases a DNA-binding dye once it is cleaved by active caspase-3/7. The fraction of caspase-3/7 positive cells was determined by calculating the ratio of apoptotic cell confluency as estimated from the fluorescent images and total cell confluency as estimated from the phase contrast images. An apoptotic index was then calculated by normalising these fractions to the value measured for 4OHT-induced sustained ERK signalling 72 hours post treatment. In accordance with flow cytometry analysis it was once more confirmed that only sustained ERK signalling is sufficient to induce apoptosis

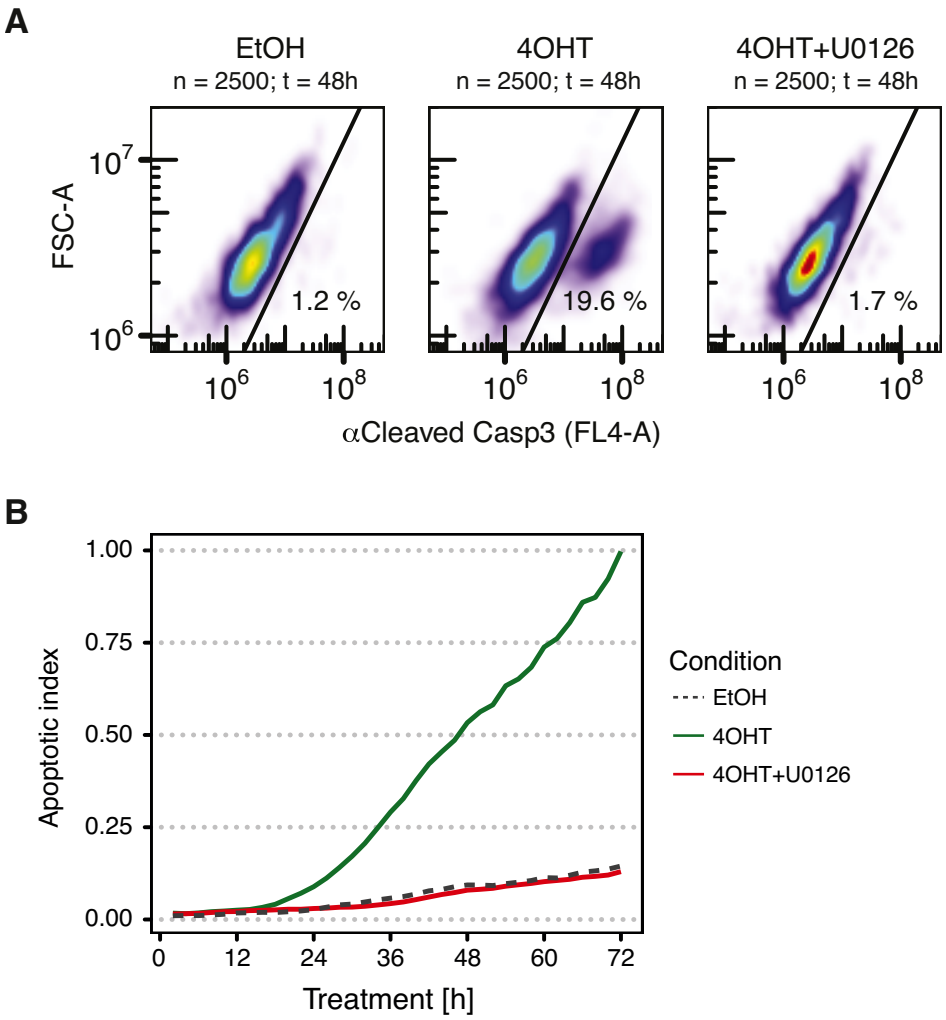


Figure 4.8 Apoptosis in sustained versus transiently induced HEK293ΔRAF1:ER cells.
A: FACS data to detect cleaved Casp3 positive cells among untreated or treated HEK293ΔRAF1:ER cells as a marker for apoptosis. EtOH: no ERK signalling. 4OHT: sustained ERK signalling. 4OHT+U0126: two-hour pulse ERK signalling.
B: Apoptotic index of HEK293ΔRAF1:ER cells upon different ERK signal durations. Treatment conditions correspond to treatments in **B**. Apoptotic index was measured with help of IncuCyte caspase-3/7 green apoptosis assay reagent in a live-cell imaging experiment for a duration of 72 hours post treatment (cf. sec. A.9).

in the tested model system, as the apoptotic index in transiently activated cells was nearly identical with the one measured for EtOH-treated control cells.

Taken together, it can be concluded that ERK signalling dynamics are successfully decoded in HEK293ΔRAF1:ER cells and translated into a duration-dependent cellular phenotype. However, it remains to be answered whether the identified potential of ILGs to serve as a duration-to-amplitude decoder in ERK signalling is a mere whim of nature, or if the module

term.name	term.id	term.size	Score	IEG	DEG	ILG
positive regulation of transcription from RNA polymerase II promoter	GO:0045944	995	3.014	9/21	2/54	5/27
positive regulation of apoptotic process	GO:0043065	569	2.772	3/21	4/54	8/27
negative regulation of protein metabolic process	GO:0051248	957	2.611	4/21	13/54	5/27
learning or memory	GO:0007611	214	2.204	5/21	0/54	3/27
positive regulation of cell proliferation	GO:0008284	814	1.618	4/21	5/54	8/27
positive regulation of biosynthetic process	GO:0009891	1749	1.59	9/21	4/54	11/27
response to lipid	GO:0033993	788	1.532	7/21	6/54	5/27
negative regulation of protein kinase activity	GO:0006469	204	1.421	2/21	6/54	1/27
negative regulation of apoptotic process	GO:0043066	844	1.338	7/21	6/54	5/27

Figure 4.9 Gene Ontology enrichment for gene expression modules controlled by ERK signalling.
Gene Ontology Biological Process (GO:BP) term enrichment for IEGs, DEGs and ILGs. Score corresponds to significant enrichment in respective cluster, where Score $= -log_{10}$(p-value). Significant enrichments are highlighted with coloured border. Background colour intensities correspond to denoted fractions and are normalised column-wise.

can functionally be linked to the observed phenotypic response. In fact, only if evidence for the latter case could be presented, one might truly consider ILGs as a receiver or decoder module in sense of Shannon's information theory. Irrespective of this, it is generally reasonable to describe a newly proposed gene expression module not only in terms of its dynamic properties, but to also examine its ultimate biological function. In this regard, previously identified IEGs and DEGs have been described to be remarkably distinct in function. Whereas IEGs have been identified to function as feed-forward elements predominantly encoding for transcription factors (boosting the expression of DEGs and inducing the expression of SRGs), DEGs have been described as a module of negative feedback regulators (Amit et al. 2007; Tullai et al. 2007; Kholodenko, Hancock, and Kolch 2010; Avraham and Yarden 2011). These negative feedback regulators include phosphatases (DUSPs), which inactivate MAP kinases (Fritsche-Guenther et al. 2011; Kidger and Keyse 2016); RNA binding proteins, which mediate degradation of IEGs (e.g. ZFP36, which binds FOS mRNA, Mukherjee et al. 2014); and other negative feedback elements, such as tumour suppressors (Amit et al. 2007).

As discussed earlier, many previously described IEGs and DEGs were also found to be regulated by ERK signalling in HEK293ΔRAF1:ER cells, suggesting that the model-based classification of PRG sub classes introduced in my thesis reliably allows to reproduce temporal cluster assignments. To further demonstrate applicability of the proposed rationale,

a functional assessment of IEGs and DEGs identified in HEK293ΔRAF1:ER cells was performed with help of gene ontology (GO) term enrichment analysis (fig. 4.9). Markedly, in accordance with the literature, IEGs were enriched for positive regulators of transcription from PolII promoter (9 out of 21: *CYR61, EGR1, EGR2, ETV5, FOS, FOSB, INSIG1, JUNB, RBM14*). Among induced DEGs, aforementioned RNA binder *ZFP36*, negative receptor feedback elements *ERRFI1* and *SPRY2*, and other negative feedback regulators of protein kinase activity were identified (*GADD45A, GADD45B, CDKN1A, TNFAIP3*). Moreover, the top upregulated DEG was tumour suppressor Tissue Factor Pathway Inhibitor 2 (*TFPI2*). Together, these results clearly reflected the previously reported positive feed forward characteristic of IEGs and the negative feedback characteristic of DEGs.

Having confirmed the different functional roles of IEGs and DEGs, in a next step, the newly defined gene cluster of ILGs was functionally characterised. Strikingly, GO term enrichment analysis suggested a distinct role of ILGs in positive regulation of apoptosis, putatively opposing involvement of IEGs in negative regulation of apoptosis. This finding suggests that the capability of ILGs to decode ERK signal duration might serve as a potential fail-save mechanism to control aberrant ERK signalling, as these positive regulators of apoptosis only come into play, when ERK is activated in a prolonged fashion. In general, it has been shown that RAF-MEK-ERK signalling is involved in positive and negative regulation of both intrinsic (mitochondrial) and extrinsic (receptor) pathway of apoptosis (Thiel, Ekici, and Rössler 2009; Cagnol and Chambard 2010). Induction of mitochondrial apoptosis pathway involves caspase-9 activation, whereas extrinsic apoptosis pathway is triggered by tumour necrosis factors (e.g. TNFα or FasL) binding to death domain receptors, in turn causing subsequent activation of caspase-8 (Fulda and Debatin 2006). Interestingly, both anti-apoptotic effects of RAF-MEK-ERK signalling (Erhardt, Schremser, and Cooper 1999; Lehmann et al. 2000; Schulze et al. 2001; Thiel, Ekici, and Rössler 2009), and pro-apoptotic effects (Wang, Martindale, and Holbrook 2000; Zhuang and Schnellmann 2006; Martin and Pognonec 2010; Cagnol and Chambard 2010; Teixeiro and Daniels 2010; Subramaniam and Unsicker 2010) have been reported for several cellular contexts. However, pro-apoptotic effects were more often reported in lymphocytes and cells of neuronal origin (Cagnol and Chambard 2010). This is remarkable since HEK293 cells have been identified as of neuronal origin (Shaw et al. 2002).

As mentioned, HEK293ΔRAF1:ER cells in particular undergo caspase-8-mediated apoptosis upon constant activation with 4OHT. However, the regulatory mechanism controlling caspase-8 activity in these cells remains to be determined (Cagnol, Obberghen-Schilling, and Chambard 2006). As it was shown that caspase-8 activation in these cells is independent of Fas-associated death domain (FADD) signalling (Cagnol, Obberghen-Schilling, and

Chambard 2006), it was later speculated that the observed caspase-8 activation might be regulated via genes from the TNF receptor super family (Cagnol and Chambard 2010). In the presented data, *TNFRSF12A* was identified as an upregulated DEG with model-derived transcriptional delay of 46 minutes and mRNA half-life of 183 minutes. Independent of its delay, its long half-life potentially enables it to translate ERK signal duration into response amplitude. It was therefore speculated that mRNA upregulation of *TNFRSF12A* could account for apoptosis in HEK293ΔRAF1:ER cells exposed to prolonged ERK activation.

To conclude this chapter, it was demonstrated that the newly identified gene expression module of immediate-late genes (ILGs) is capable of translating ERK signal duration into response amplitude. It was shown that both gene expression timing and the gene-specific capacity to decode signal duration are linked to mRNA half-life. The overall significance of these findings was validated in different model systems for ERK signal duration, namely PC12 and MCF7 cells. Lastly, based on functional analyses it was suggested that ILGs might serve as a ERK signal duration receiver module potentially providing a fail-save mechanism to control aberrant ERK signalling.

5 Conclusion and discussion

The idea of mRNA half-life being important for the kinetics of gene induction is as old as the discovery of the messenger RNA itself (Jacob and Monod 1961). Yet only the advent of high-throughput technologies allowed to test this hypothesis in a genome-wide manner. Since that time several studies have specifically demonstrated that short-lived transcripts respond early and that long-lived transcripts respond late to external stimuli. For example, the temporal order of gene expression was shown to be governed by mRNA half-life upon H2O2-induced stress in yeast (Shalem et al. 2008), upon NF-kB signalling in mouse 3T3 fibroblasts (Hao and Baltimore 2009) and upon IL-2 signalling in murine T cells (Elkon et al. 2010). In my thesis, I now demonstrated that mRNA half-life likewise governs gene expression timing upon ERK signalling in HEK293 cells. More importantly though, my analysis indicated a second immanent role for mRNA half-life, namely facilitating the decoding of ERK signal duration. In an integrative approach, combining experimental and computational methods, I presented evidence for a new, distinct module of primary response genes which employs this dual function of mRNA half-life. Paradoxically, these genes share properties with both IEGs and DEGs. Accordingly, they were termed immediate–late genes (ILGs), as they are induced immediately like IEGs, but respond late like DEGs. In the last part of my thesis, I will now briefly summarise the analyses that allowed me to arrive at these conclusions and thoroughly discuss their limitations and underlying assumptions.

The first part of my thesis discussed the importance of an appropriate model system to study gene expression timing and ERK signal duration decoding. Here, it was argued that the gene regulatory effects of ERK signalling are potentially blurred by the complex interconnected nature of signalling networks, when classical growth factor–induced model systems are studied. Instead, it was stated that more controllable systems should be preferred which allow to short-cut signalling and to relate signalling input dynamics to gene expression output dynamics. In case of the employed experimental model system, this involved direct activation at the level of RAF, rendering feedback loops moot. Overall, even though more complex signal dynamics could certainly have been tested with an optogenetic approach (c.f. below discussion of Johnson, Shvartsman, and Toettcher 2018), the presented insights gained from the HEK293ΔRAF1:ER model system clearly demonstrated its excellent suitability to investigate ERK signal duration decoding.

A thorough analysis of the data obtained from the experimental system was achieved by utilising a simple mathematical model of gene expression. In summary, four major insights were drawn directly from or with help of the mathematical framework that should be mentioned here. First, *in silico* simulations of gene expression changes upon different ERK signalling scenarios suggested the above mentioned dual role for mRNA half-life that was later confirmed experimentally, which is to govern gene expression timing and to facilitate signal duration decoding. Secondly, the framework provided a theoretical rationale to identify different temporal gene expression modules responding to ERK signalling. Thereby, the novel gene expression module of ILGs was identified and characterised by its distinct parameter properties. Thirdly, the generalisability of the framework implicated that combined information about phospho-ERK levels and gene-specific rate constants is sufficient to explain the gene regulatory dynamics in response to ERK signalling. And lastly, the analysis once more underlined that gene expression parameters are continuous. Hence, it should be noted again to view the proposed temporal gene cluster definitions as a heuristic to aid interpretation and to ease comparison of results across literature.

Considering the modelling framework in general, it is somewhat surprising that a simple one-step model of gene expression as employed here described the obtained data with reasonable accuracy and was successfully adopted to semi-quantitatively predict gene expression dynamics upon different signalling scenarios. Most significantly, the model assumes that mRNA degradation rates are constant; it combines transcription and RNA processing into a single step of mRNA synthesis; and it models any potential multi-step processes that may need to take place before ERK signalling–mediated transcription can start with help of a simple delay parameter. In contrast, over the last years, several more complex mathematical frameworks have been presented to analyse the underlying regulatory aspects of mRNA dynamics (Zeisel et al. 2011; Rabani et al. 2014; Pretis et al. 2015; Cheng et al. 2017). Interestingly, although most of these studies allowed for dynamic changes in degradation rates and all of them employed two-step or multi-step models to either disentangle the different contributions of transcription and processing rates or to model sequential transcription factor logics, a number of conclusions arise from these studies which, taken together, can justify the use of a more simple one-step model instead. First, the studies demonstrated that changes in mRNA expression are mainly governed by changes in mRNA transcription, whereas processing and degradation rates were only altered for a minority of regulated genes (4 % and 10 % of genes respectively, Rabani et al. 2014). Secondly, average precursor half-lives were identified to be about one order of magnitude smaller than average half-lives of processed mRNAs (Rabani et al. 2014; Alpert, Herzel, and Neugebauer 2017), suggesting that induction kinetics are not dominated by precursor dynamics. Hence, one could indeed

deduce from these findings that a simple one-step model serves as a reasonable abstraction to genuinely describe mRNA dynamics.

To be fair, however, a more simple model is necessarily accompanied by a number of limitations. Most importantly in this case, the inability of assessing processing rates demands some further discussion. Although mRNA induction kinetics may globally not notably be affected by processing rates, the timing of particular mRNAs might still be governed by dynamic changes in splicing rates. As a brief reminder, previous studies had shown that late induction of genes can indeed be caused by both restrained splicing of pre-mRNA (Zeisel et al. 2011; Hao and Baltimore 2013; Rabani et al. 2014; Feldman and Yarden 2014) or long mRNA half-lives (Yang et al. 2003; Shalem et al. 2008; Hao and Baltimore 2009; Elkon et al. 2010; Nagashima et al. 2015; Porter, Fisher, and Batchelor 2016; Cheng et al. 2017). In this regard, the evidence presented in my thesis for long mRNA half-lives to serve as a key determinant for late gene induction of primary response genes upon ERK signalling partly needs to be considered in light of this limitation. By design, a potential additional contribution of restrained splicing could not be tested, given the employed framework and the obtained data. For this, either different microarrays probing intronic sequences (Zeisel et al. 2011) or a different sequencing library preparation protocol (ribosomal RNA depletion instead of poly-A enrichment) would have been required to also allow quantification of precursor mRNA levels, as other studies have demonstrated (Rabani et al. 2014; Pretis et al. 2015). However, juxtaposition of mRNA induction kinetics and the results gained from transcriptional shut-down and metabolic labelling experiments clearly underlined the vast impact of mRNA half-life on gene expression timing. As an outlook, the additional assessment of processing rates could provide valuable information on how restrained splicing may additionally contribute to late induction of primary response genes in HEK293 cells and to what extent the current model framework might have inadvertently already captured this effect in form of the transcriptional delay parameter (which in this case would need to be reconsidered as a restrained processing parameter).

Apart from assuming zero-order mRNA synthesis, an additional but rather minor limitation of the model is given by the fact that it assumes first-order mRNA decay. Whereas most studies examining mRNA dynamics assume mRNA degradation to be a first-order process, non-exponential decay —i.e. age-dependent degradation of mRNA— has been reported (Deneke, Lipowsky, and Valleriani 2013a; Deneke, Lipowsky, and Valleriani 2013b), but could not be examined with the presented model. With regard to mRNA half-life, it was also noted that model fits of gene induction kinetics yielded estimates for IEGs and ILGs with reasonable accuracy, but systematically underestimated mRNA half-lives of DEGs when compared to ActD- and 4SU-based measurements. For these genes, the introduced de-

lay parameter Δt potentially obscured a meaningful biological interpretation of the modelled degradation rate. However, this limitation was justified by a number of advantages provided by this additional parameter. Most obviously, it allowed to quantify transcriptional delays and thus their contribution to gene expression timing of late primary response genes. Furthermore, it also provided an additional degree of freedom to the basic model of gene expression and thereby rendered it to mimic a multi-step induction process, but without the need for solving two or more coupled ordinary differential equations.

Considering the obtained transcriptome data in general, one must also not forget to point out that mRNA dynamics were assessed from bulk data, i.e. averaged parameter values were estimated for a population of cells. As single-cell experiments using reporter constructs have demonstrated that transcription has a stochastic component and takes place in bursts (Elowitz et al. 2002; Suter et al. 2011; Munsky, Neuert, and Oudenaarden 2012; Senecal et al. 2014), it should be noted that any such potential cell-by-cell variability is masked in bulk measurements. Thus, different biological interpretations of an immediate but slow mRNA accumulation, as observed for ILGs, are possible. On the one hand such dynamics could indeed be caused by a corresponding gradual accumulation in each individual cell, but on the other hand they might likewise reflect the cumulative result of a switch-like induction with a highly variable transcriptional delay (Purvis and Lahav 2013). This particular ambiguity could not be solved by the modelling of bulk gene induction kinetics, but required consideration of the additionally obtained, independent measurements of nascent mRNA production and mRNA degradation. Here, again analyses of 4SU and ActD data suggested that the immediate but slow accumulation of ILG mRNAs could indeed be attributed to their immediate transcription and mRNA longevity. In this regard, single molecule imaging approaches to track RNA expression *in situ* (smFISH) would provide additional information on cell-by-cell variability of ILG induction kinetics and complement the available data on IEGs (Senecal et al. 2014; Battich, Stoeger, and Pelkmans 2015; Gómez-Schiavon et al. 2017).

Having identified mRNA half-life as an important dynamic feature to enable late induction of primary response genes, the second main question to be answered was if and how primary response genes can function as molecular decoders to interpret ERK signal duration. As elaborated in the general introduction, previous studies had often focused on stimulus encoding, but comparatively few attempts have been made to study signal decoding (cf. sec. 1.3). The FOS protein sensor model forms an exception and serves as an example for signal duration decoding in 3T3 cells and PC12 cells (Murphy et al. 2002; Pellegrino and Stork 2006). Yet, it remained unclear whether or not these cells possess additional mechanisms of signal duration decoding and if the principle of duration-to-amplitude decoding initially proposed for

PRG protein products could also be identified in PRG mRNAs and hence be adapted to postulate a mRNA sensor/decoder model. As an initial effort in this regard, feed-forward regulation of mRNA stability had been suggested to enable ERK signal duration decoding in MCF7 cells (Nagashima et al. 2015). Here, mRNA half-lives were estimated based on gene-induction kinetics and experimentally validated for a small subset of genes. Analogous to the FOS protein sensor model, the authors linked duration decoding to feed-forward regulation of mRNA stability. Recently, the idea that mRNA half-life is involved in signal decoding has also been shown for signal frequency decoding of p53 signalling (Porter, Fisher, and Batchelor 2016). Here, short-lived transcripts relay the oscillatory pattern of p53 signalling pulses to response pulses, whereas only long-lived transcripts decode the pulses by translating them into response amplitude. In my thesis, I now demonstrated that this principle also transfers to decoding of signal duration in ERK signalling and that the identified decoders do not require any additional regulation, although previously suggested (Nagashima et al. 2015). Similar to the reported example of p53 signal frequency decoding, mRNA longevity intrinsically enables mRNAs to decode ERK signal duration in HEK293 cells. Although additional feed-forward regulation of mRNA stability might potentially increase signal duration decoding capacity of primary response genes, mRNA longevity–enabled interpretation of ERK signal duration was identified as a general principle of duration-to-amplitude decoding. Moreover, the newly identified gene expression module of ILGs employs this principle and thus serves as an example for the proposed mRNA sensor/decoder model.

Generally, the identification of decoder elements is of particular interest, as the decoder ultimately determines which signalling feature is a functionally relevant coding feature (Behar and Hoffmann 2010). For example, to get from structural genes to proteins, tRNAs function as decoders to translate the genetic code into an amino-acid sequence. Hence, one could say that tRNAs determine nucleotide triplets to be the functionally relevant coding feature of the genetic code. In case of ERK signal duration decoding, it was recently questioned whether it is the actual persistence of the signal or rather its cumulative load which is interpreted by molecular decoder elements (Ryu et al. 2015; Gillies et al. 2017; Johnson, Shvartsman, and Toettcher 2018). To discriminate between a persistence sensor model and a cumulative load sensor model, one study used an optogenetic system to dynamically control ERK signalling and compared short but repeated signalling pulses with one long pulse of the same overall activity duration. Markedly, it was found that the profiled phenotype (in this case ectopic contraction in early Drosophila embryos) is triggered by the cumulative load and not by the persistence of the sustained signal (Johnson, Shvartsman, and Toettcher 2018). In accordance, other studies had previously reported that linear integration of ERK activity predominates over persistence detection (Gillies et al. 2017) and that ERK activity

frequency modulation by synthetic multi-pulse regimes can recapitulate effects of sustained ERK signalling and rewire cell fate in PC12 cells (Ryu et al. 2015). Following the idea that the decoder determines the signalling code, the parametric properties of ILGs likewise favour a cumulative load sensor model. In contrast, the earlier proposed notion of feed-forward regulated mRNA decoder elements (Nagashima et al. 2015) would have rather suggested predominance of ERK activity persistence detection.

Identification of ILGs as molecular decoders of ERK signal duration was based on a crucial distinction between the actual decoding of duration and the mere transmission of signalling dynamics. In this regard, it was demonstrated that both short-lived and long-lived transcripts can discriminate different signal durations. However, only long-lived genes truly decode signal duration by translating it to response amplitude (mRNA decoder model), whereas short-lived genes relay signal duration to response duration, postponing the task of decoding (mRNA relay model). When considering protein dynamics in addition, a total of four different scenarios by which signal duration could translate to the level of mRNA and protein dynamics were identified. In addition to highlighting these qualitative differences in the interpretation of ERK signal duration, a likewise important aspect was to address the relation between ERK signal duration decoding and mRNA half-life in a more quantitative manner. First of all, the continuous nature of gene expression implies that different genes decode duration to varying extents. However, so far, the capacity of ERK signalling–regulated mRNAs to decode duration had not yet been quantified. For this purpose, gene-wise signal duration decoding capacities (SDDC) were quantified with help of duration-to-amplitude profiles which in turn were generated from sampling induction kinetics over a range of different ERK signal durations in HEK293ΔRAF1:ER cells. Markedly, the obtained estimates allowed to identify a positive correlation between the mRNA half-life and the duration decoding capacity of genes responding to ERK signalling.

Evidently, a final important analysis addressed functional implications and the overall relevance of the presented findings. Here, it was found that the principle of signal duration being translated into response amplitude is conserved in rat PC12 cells and human MCF7 cells, two cell systems which serve as paradigm models for cell fate decisions based on signal duration. In general, mRNA half-life is a strong predictor for response dynamics in these systems. With regard to functional implications, gene term enrichment analysis confirmed previous attributions for IEGs and DEGs and suggested a distinct function for ILGs. More precisely, genes encoding for transcription factors were enriched among the short-lived, fast responding IEGs and negative feedback regulators were enriched among the delayed responding DEGs. Moreover, many ILGs and some additional long-lived DEGs encoded for genes that are involved in positive regulation of apoptosis, which is the cell fate for sustained ERK signalling

in HEK293 cells, suggesting that mRNA longevity–enabled decoding of signal duration by these genes might ultimately account for the observed cellular response.

With regard to the overall relevance of the presented findings, it should finally be noted once again that the used synthetic cell culture system mimics the activation of the oncogene RAF. Oncogenic hyperactivation of RAF1 / BRAF is a pro-survival signal in many contexts. However, many cell types activate fall-back programs to oppose the overactive signalling of RAF. Like DEGs, the newly identified cluster of ILGs may serve as such a fall-back module to counteract pro-survival signals sent out by sustained RAF activation. When benign tumours progress into malignant ones many negative feedback mechanisms that conferred robustness before are lost (Friday et al. 2008). A deep understanding of feedback modules or fail-safe mechanisms in the cluster of ILGs that decode sustained oncogenic signalling is therefore crucial to better understand what distinguishes proto-oncogenic from oncogenic signalling.

Appendix

A Materials and methods

Throughout the main text of my thesis, I tried to describe both experimental and computational methods along with the results they facilitated to an extent that would aid interpretation but not disrupt the reading flow. Hence, this chapter now provides supplementary information on the employed methods which did not find its way into the main text.

A.1 ERK signalling network alterations across cancers

Mutation data, copy number alteration (CNA) data and patient survival data was obtained from cBioPortal using the application programming interface (API) provided by the platform (Gao et al. 2013). Mutation and CNA data was obtained for two of the curated subsets of genes provided by cBioPortal, namely *Ras-RAF-MEK-Erk/JNK signaling* (26 genes) and *RTK signaling family* (16 genes). The analysis was performed on cBioPortal snapshot v1.17.1 in October 2018. At that time, 233 studies across 96 cancer entities with a total of 40247 unique cases were available to query via cBioPortal. For a subset of 16830 cases, both mutation and CNA data was available. Here, KRAS, BRAF and EGFR were found to be the top most frequently altered genes across all cases. They were found to be particularly enriched in 8 different cancer entities (at a frequency greater than 25% for large cohorts with at least 500 patients or at a frequency greater than 50% for small cohorts with at least 50 patients). A subset of 41 studies entailing 8338 unique cases was available for these 8 cancer entities and considered for a meta-analysis to calculate mutation and amplification frequencies, as well as alteration-specific survival rates. Kaplan-Meier survival curves where estimated using R package survminer (Kassambara, Kosinski, and Biecek 2017). Summary statistics for all considered cancer types and studies including references are provided in supplementary tables B.1 and B.2.

A.2 Cell culture, microarray hybridization and phosphoprotein assay

HEK293ΔRAF1:ER cells (Samuels et al. 1993; reviewed in McMahon 2001) were cultured in complete DMEM high glucose without phenol red with 10% fetal calf serum supplemented

with antibiotics (pen/strep). Before stimulation, cells were starved in serum free medium over-night. Cells were stimulated with 4-hydroxy Tamoxifen (Sigma-Aldrich H7904; 0.5 μM), U0126 (20 μM), EGF (25 ng/ml), FGF1 (50 ng/ml), or IGF (100 ng/ml). Translation was inhibited with Cycloheximide (10 μM), and transcription was inhibited with Actinomycin D (5 μM). RNA for microarray hybridization was isolated with TRIZOL reagent. cDNA was fragmented, labelled and hybridized to Affymetrix Human Gene 1.0 ST Arrays. Phosphoprotein levels were assessed with Bio-Plex (Bio-Rad) as described previously (Klinger et al. 2013). Metabolic labelling of RNA with 200 μM 4SU one hour before harvesting followed by RNA-seq was performed as described previously (Schueler et al. 2014).

A.3 Identification of differentially expressed genes from microarray data

Fluorescence intensities from scanned microarrays were processed and analysed in R. Background correction, quantile normalisation, probe set summarisation and $\log_2$ transformation was performed with help of robust multichip average algorithm (RMA) (Irizarry 2003). Probe sets were annotated with R package hugene10sttranscriptcluster.db. All probe sets mapping to a HUGO symbol identifier were considered. For transcripts represented by multiple probe sets, the probe set with highest mean expression across samples was considered. Transcripts expressed below median expression in all samples were excluded from analysis. Log_2 fold changes were calculated with respect to mean expression in untreated samples (UT-1 and UT-3). UT-2 was excluded due to strong dissimilarity to UT-1 and UT-3 in cluster analysis of correlation values and putative contamination. Log_2 fold changes for independently obtained EGF, and FGF time course data were calculated with respect to mean expression in corresponding untreated samples (UT-1-n, UT-2-n, UT-3-n). To account for expression level-dependent variations, an empirical null model was based on replicates for two-hour 4OHT treatment. For this, transcripts were ranked by their mean expression across replicates and a moving average with window size k = 2000 was calculated to serve as an expected variance measure for a given expression level. Z-scores for each transcript p_i in each sample j were calculated accordingly:

$$z_{i,j} = \frac{p_{i,j} - p_{i,UT}}{\langle sd\left(\frac{p_{i,j}-p_{i,UT}}{2}\right)\rangle} \tag{A.1}$$

Genes exceeding an absolute z-score of 5.6 in 4OHT time course data were considered regulated (1490 up-, 2037 downregulated). This corresponded to an average false discovery rate (FDR) of 1% in 4OHT time course data. Here, false positives were estimated by counting

transcripts detected differentially expressed between one replicate and the mean of the two other replicates of the two-hour 4OHT treatment samples. For all downstream analyses 4OHT-regulated genes were further filtered in two steps. First, regulated genes were tested against a random set of unregulated genes (of the same size) for their $\log_2$ fold change standard deviation ($SD\log_2fc$) across all samples. This was done to filter out a large fraction of erroneously detected genes, which were unaltered across all samples when the untreated condition was left out. Here, a $SD\log_2fc$ cut-off was defined at FDR of 5% (253 up-, 234 downregulated genes remained). Secondly, genes induced in a non-monotonic fashion that could not be fitted to the employed one-step model were excluded from the analysis. All remaining genes (189 up-, 146 downregulated) are referred to as differentially expressed in the main text. Differentially expressed genes were checked for significant z-scores in CYHX samples. Differentially expressed genes also significantly induced in any sample of parallel CYHX treatment were considered primary response genes (PRGs).

For the parallel assessment of publicly available microarray time-course data, raw data on PC12 (Offermann et al. 2016) and MCF7 (Saeki et al. 2009) was downloaded from gene expression omnibus (GEO). Data was then preprocessed and analysed analogously to HEK293ΔRAF1:ER microarray data.

A.4 RNA-sequencing data generation and preprocessing

Total RNA was extracted with Trizol. For labelling experiments, 4SU-labelled and unlabelled fractions were separated as described previously (Baltz et al. 2012). Sequencing libraries were prepared using Illumina TruSeq mRNA Library Prep Kit v2 and sequenced on Illumina HiSeq 2000. Read files were demultiplexed and sequencing adapters were trimmed using flexbar (Dodt et al. 2012). Reads were mapped with STAR aligner v2.4.1 (Dobin et al. 2013) on hg19 using GENCODE v19 for annotation and counted with subread featureCounts (Liao, Smyth, and Shi 2014). Raw read counts were normalised with edgeR TMM (Robinson, McCarthy, and Smyth 2010) and analysed with R package DTA (Miller et al. 2011). TPM values were calculated by normalising read counts to transcript length and to number of sequenced reads and scaled with a factor of 10^6 (Li and Dewey 2011). The entire preprocessing pipeline was written in Snakemake (Köster and Rahmann 2012).

A.5 Modelling of mRNA dynamics and calculation of response amplitudes

Gene expression data was fitted to basic and extended model for mRNA dynamics as described in the main text using Nelder-Mead method implemented in R package optimx. The empirical null described above was again considered to account for expression level dependent variance. Residuals were weighted with respect to the expected variance for a given expression value. Optimised model fits for the basic and extended model were compared using a log-likelihood test (reviewed in Bentler and Bonett 1980).

For a given expression of a gene at time t, response amplitude was deduced from gene-wise parameter estimates of k_0, k and γ:

$$response\ amplitude\ (t) = \left(expression\,(t)\ -\ \frac{k_0}{\gamma}\right)\ /\ \left(\frac{k_0 + k}{\gamma} -\ \frac{k_0}{\gamma}\right) \tag{A.2}$$

To obtain semi-quantitative $\log_2$ fold change predictions for growth factor-induced gene expression, gene-wise fitted model parameters and input functions for ERK dependent transcription rate k were fed to the complete model.

A.6 ActD- and 4SU-based mRNA half-life estimation

Half-life estimates based on ActD-mediated transcriptional shut-down were derived from microarray gene expression time course data (Fig 2.3A). Since quantile normalisation assumes constant total RNA levels across samples (Bar-Joseph, Gitter, and Simon 2012), RMA was performed without quantile normalisation for ActD samples. Samples were instead normalised to median expression of 61 long-lived mRNAs ($t_{1/2} > 16$ hours) consistently identified in two published data sets on human mRNA half-life (Friedel et al. 2009; Yang et al. 2003). Both time series were than fitted to an exponential decay function of form $M\,(t) = M_0 + e^{-\gamma t}$ to infer decay rates γ.

Half-life estimates based on metabolic labelling with 4SU followed by RNA sequencing were calculated using all three fractions of RNA, i.e. total RNA, labelled RNA (eluate) and unlabelled RNA (flow-through). Dynamic transcriptome analysis (DTA) was used for quantification (Miller et al. 2011). Median, mean and standard deviation of half-lives for all expressed genes were calculated from the three different data sets (ActD ON, ActD OFF, 4SU).

A.7 Signal duration decoding capacity

Signal duration decoding capacity (SDDC) was estimated from duration-to-amplitude profiles. For this, sampled ERK signal durations were normalised to the longest duration tested, resulting in relative signal durations between 12.5% (corresponding to 30 minutes of ERK activity) and 100% (corresponding to 8 hours of ERK activity). For a given gene i, SDDC was defined as one minus the area under the curve (AUC, using linear interpolation between samples):

$$SDDC_i = 1 - AUC_i \tag{A.3}$$

A.8 qRT-PCR primers and western blot antibodies

cDNA was synthesised using High-Capacity RNA-to-cDNA Kit (Applied Biosystems #4387406). qRT-PCR was performed using Taqman gene expression assay (Thermo Fisher #4304437) with following Taqman primers (Thermo Fisher):

#	Gene	Primer	#	Gene	Primer
1	ARC	Hs01045540_g1	10	JUNB	Hs00357891_s1
2	CLU	Hs00156548_m1	11	NR4A1	Hs00374226_m1
3	DUSP1	Hs00610256 _g1	12	PGK1	Hs00943178_g1
4	DUSP6	Hs01044001_m1	13	PPP1R15A	Hs00169585_m1
5	EGR1	Hs00152928_m1	14	PTGS2	Hs00153133_m1
6	EGR2	Hs00166165_m1	15	TFPI2	Hs04334126 _m1
7	FOS	Hs00170630_m1	16	TNFRSF12A	Hs00959047_g1
8	FOSB	Hs00171851_m1	17	ZCCHC12	Hs00381614_m1
9	FOSL1	Hs04187685_m1	18	ZFP36	Hs00185658_m1

Protein was extracted using Bio-Rad Cell Lysis Buffer (#171-304006M). Concentration was determined using Thermo Fisher Pierce BCA Protein Assay (#23228). 25-50 μg of purified protein was used for blotting. Images were acquired using LI-COR Odyssey Scanner. Western blot antibodies: EGR1 (Santa Cruz sc-110), FOS (Cell Signaling #2250), CLU (Santa Cruz sc-8354), FOSL1 (Santa Cruz sc-376148).

A.9 Flow cytometry and live cell imaging

For flow cytometry and live cell imaging, cells were treated with EtOH, 4OHT or 4OHT plus U0126 (two hours post 4OHT treatment). For flow cytometry, cells were harvested 48h after treatment and fixated in 2% paraformaldehyde (PFA) for 10 minutes at RT. Cells were permeabilised in methanol and incubated on ice for 30 minutes. Immunostaining was performed by incubating cells with Cleaved Caspase-3 rabbit mAb (Cell Signaling #4440) for one hour. For live cell imaging, cells were additionally treated with caspase-3/7 green apoptosis assay reagent (Essen BioScience #4440) and tracked with IncuCyte ZOOM device for a total of 72 hours. An apoptotic index was calculated by assessing the ratio of cell confluency in fluorescent images (representing caspase-3/7 positive cells per area) and cell confluency in phase contrast images (representing all cells per area) and by then normalising this fraction to the maximum value measured across treatments.

A.10 Data availability

Primary data Both microarray gene expression data and metabolic labelling RNA-seq data presented in my thesis are accessible from gene expression omnibus (GEO) under SuperSeries accession number GSE93611.

Referenced data Yang et al. (2003) and Friedel et al. (2009) were considered for mRNA half-life estimates. Offermann et al. (2016, GSE74327) and Saeki et al. (2009, GSE13009) were considered for PC12 and MCF7 gene expression time course data.

B Supplementary tables

B.1 Considered cancer types to assess ERK signalling network alterations

Table B.1 Considered cancer types to assess ERK signalling network alterations. List of eight different cancer types considered in fig. 1.2. The full analysis is described in sec. A.1. As some studies share cases (e.g. TCGA original and TCGA provisional releases), the total number of considered cases for each cancer type may not match the sum of cases per study shown in table B.2. MUT: Cases with mutation data. CNA: Cases with copy number alteration data. Clin.: Cases with clinical data.

#	Cancer	Studies	Cases	MUT	CNA	Clin.
1	Ampullary Cancer	1	160	160	0	159
2	Colorectal Cancer	8	2583	2560	1751	1627
3	Glioma	4	598	449	580	589
4	NSCLC	10	2654	2110	860	983
5	Pancreatic Cancer	5	776	772	293	185
6	Pilocytic Astrocytoma	1	96	96	0	0
7	Melanoma	8	844	842	433	541
8	Thyroid Cancer	4	627	578	620	625

B.2 Considered cancer studies to assess ERK signalling network alterations

Table B.2 Considered cancer studies to assess ERK signalling network alterations. PubMed references for the 41 different cancer studies downloaded from cBioPortal (v1.18.0) and considered in fig. 1.2.

#	Cancer	Reference	PubMed	Cases
1	Ampullary Cancer	Gingras et al., Cell Reports 2016	26804919	160

Continued on next page.

Table B.2 Considered cancer studies.

#	Cancer	Reference	PubMed	Cases
2	Colorectal Cancer	TCGA, PanCancer Atlas, Cell 2018	29625048	439
3	Colorectal Cancer	Giannakis et al., Cell Reports 2016	27149842	619
4	Colorectal Cancer	Seshagiri et al., Nature 2012	22895193	72
5	Colorectal Cancer	Brannon et al., Genome Biol 2014	25164765	138
6	Colorectal Cancer	TCGA, Provisional		619
7	Colorectal Cancer	TCGA, Nature 2012	22810696	269
8	Colorectal Cancer	Yaeger et al., Cancer Cell 2018	29316426	1134
9	Colorectal Cancer	TCGA, PanCancer Atlas, Cell 2018	29625048	155
10	Glioma	TCGA, Provisional		594
11	Glioma	TCGA, PanCancer Atlas, Cell 2018	29625048	592
12	Glioma	TCGA, Nature 2008	18772890	199
13	Glioma	TCGA, Cell 2013	24120142	573
14	NSCLC	Imielinksi et al., Cell 2012	22980975	183
15	NSCLC	Rizvi et al., Science 2015	25765070	34
16	NSCLC	TCGA, Provisional		516
17	NSCLC	TCGA, PanCancer Atlas, Cell 2018	29625048	566
18	NSCLC	TCGA, Nature 2014	25079552	230
19	NSCLC	Ding et al., Nature 2008	18948947	162
20	NSCLC	Jordan EJ et al., Cancer Discov 2017	28336552	915
21	NSCLC	Rizvi et al., JCO 2018	29337640	240
22	NSCLC	TCGA, Nat Genet 2016	27158780	1144
23	NSCLC	VavalË et al., Lung Cancer 2017	27346245	41
24	Pancreatic Cancer	Biankin et al., Nature 2012	23103869	99
25	Pancreatic Cancer	Bailey et al., Nature 2016	26909576	383
26	Pancreatic Cancer	TCGA, Provisional		185
27	Pancreatic Cancer	TCGA, PanCancer Atlas, Cell 2018	29625048	184
28	Pancreatic Cancer	Witkiewicz et al., Nat Commun 2015	25855536	109
29	Pilocytic Astrocytoma	Jones et al., Nat Genet 2013	23817572	96
30	Melanoma	Hodis et al., Cell 2012	22817889	121
31	Melanoma	Van Allen et al., Cancer Discov. 2012	24265153	78
32	Melanoma	Berger et al., Nature 2012	22622578	25
33	Melanoma	TCGA, Provisional		448
34	Melanoma	TCGA, PanCancer Atlas, Cell 2018	29625048	444

Continued on next page.

Table B.2 Considered cancer studies.

#	Cancer	Reference	PubMed	Cases
35	Melanoma	Hugo et al., Cell 2016	26997480	38
36	Melanoma	Chakravarty et al., JCO 2017	28890946	66
37	Melanoma	Krauthammer et al., Nat Genet 2012	22842228	91
38	Thyroid Cancer	TCGA, Provisional		510
39	Thyroid Cancer	TCGA, PanCancer Atlas, Cell 2018	29625048	500
40	Thyroid Cancer	TCGA, Cell 2014	25417114	499
41	Thyroid Cancer	Landa et al., J Clin Invest 2016	26878173	117

End of table.

B.3 Gene-wise parameter estimates for IEGs, DEGs, ILGS and SRGs

Table B.3 Gene-wise parameter estimates for IEGs, DEGs, ILGS and SRGs.

Gene	Cluster	Δt	t_{1_2}	r	Log_2 fc	z	ENSEMBLE
BTG2	IEG	0	53	53	1.30	19.0	ENSG00000159388.5
CYR61	IEG	0	21	21	0.97	11.0	ENSG00000142871.11
DNAJB1	IEG	0	12	12	0.87	14.0	ENSG00000132002.3
DUSP1	IEG	0	10	10	0.98	17.0	ENSG00000120129.5
EGR1	IEG	0	29	29	4.00	72.0	ENSG00000120738.7
EGR2	IEG	0	117	117	3.50	51.0	ENSG00000122877.9
EGR3	IEG	0	68	68	4.10	48.0	ENSG00000179388.8
ETV5	IEG	0	87	87	1.40	19.0	ENSG00000244405.3
FOS	IEG	0	24	24	2.80	39.0	ENSG00000170345.5
FOSB	IEG	0	50	50	2.20	26.0	ENSG00000125740.9
IER2	IEG	0	23	23	1.10	14.0	ENSG00000160888.6
IER3	IEG	0	98	98	1.90	28.0	ENSG00000137331.11
INSIG1	IEG	0	64	64	1.10	21.0	ENSG00000186480.8
JUNB	IEG	0	81	81	2.30	33.0	ENSG00000171223.4
NAGK	IEG	0	27	27	0.36	6.6	ENSG00000124357.8
NPTX2	IEG	0	106	106	1.30	21.0	ENSG00000106236.3
RBM14	IEG	0	22	22	0.65	11.0	ENSG00000239306.4
RNF122	IEG	0	112	112	1.60	24.0	ENSG00000133874.1

Continued on next page.

Table B.3 Gene-wise parameter estimates for IEGs, DEGs, ILGS and SRGs.

Gene	Cluster	Δt	t_{1_2}	r	Log_2 fc	z	ENSEMBLE
TIMP3	IEG	0	41	41	1.00	16.0	ENSG00000100234.11
TRIB1	IEG	0	98	98	1.60	29.0	ENSG00000173334.3
YRDC	IEG	0	63	63	1.30	20.0	ENSG00000196449.3
ABHD5	DEG	93	92	185	0.60	9.1	ENSG00000011198.3
ARL5B	DEG	70	232	302	0.68	13.0	ENSG00000165997.4
ARRDC4	DEG	51	57	108	2.30	33.0	ENSG00000140450.7
BHLHE40	DEG	116	517	633	0.68	9.9	ENSG00000134107.4
C2orf42	DEG	60	121	181	0.45	5.7	ENSG00000115998.3
C6orf141	DEG	351	10	361	1.50	16.0	ENSG00000197261.7
CA2	DEG	103	27	130	0.93	17.0	ENSG00000104267.5
CDKN1A	DEG	82	175	257	1.60	23.0	ENSG00000124762.9
COQ10B	DEG	101	30	131	0.82	12.0	ENSG00000115520.4
DDX21	DEG	104	52	157	0.50	9.5	ENSG00000165732.8
EFNB2	DEG	104	48	152	0.74	11.0	ENSG00000125266.6
ERRFI1	DEG	112	10	122	0.66	9.7	ENSG00000116285.8
FAM57A	DEG	108	23	131	0.94	17.0	ENSG00000167695.10
FAM83G	DEG	117	10	127	1.20	13.0	ENSG00000188522.10
GADD45A	DEG	96	181	277	0.84	11.0	ENSG00000116717.7
GADD45B	DEG	103	24	127	2.30	33.0	ENSG00000099860.4
GEM	DEG	43	101	144	1.80	18.0	ENSG00000164949.3
GPR50	DEG	79	259	338	1.80	23.0	ENSG00000102195.7
GTF2B	DEG	99	73	173	0.55	8.7	ENSG00000137947.7
HBEGF	DEG	102	38	140	1.10	15.0	ENSG00000113070.6
HOMER1	DEG	104	156	260	1.40	21.0	ENSG00000152413.10
ID4	DEG	96	88	183	0.58	8.6	ENSG00000172201.6
KDM6B	DEG	99	152	251	1.10	14.0	ENSG00000132510.6
KIAA0020	DEG	107	95	202	0.81	9.5	ENSG00000080608.9
KLF10	DEG	54	46	100	1.40	19.0	ENSG00000155090.10
KRT8	DEG	160	570	730	1.60	25.0	ENSG00000170421.7
MAK16	DEG	109	45	154	0.54	7.6	ENSG00000198042.6
MXD1	DEG	63	33	97	0.65	9.3	ENSG00000059728.6
NDRG1	DEG	58	77	135	0.92	17.0	ENSG00000104419.10
NIPAL1	DEG	108	70	178	1.30	19.0	ENSG00000163293.7

Continued on next page.

Table B.3 Gene-wise parameter estimates for IEGs, DEGs, ILGS and SRGs.

Gene	Cluster	Δt	t_{1_2}	r	Log_2 fc	z	ENSEMBLE
NRARP	DEG	100	70	170	2.00	30.0	ENSG00000198435.2
PDK4	DEG	105	51	156	1.40	17.0	ENSG00000004799.7
PLAT	DEG	96	55	151	0.85	12.0	ENSG00000104368.13
PMAIP1	DEG	56	53	108	0.46	8.7	ENSG00000141682.11
PMM2	DEG	111	10	121	0.60	7.5	ENSG00000140650.7
PPP1R15A	DEG	30	2879	2909	1.50	22.0	ENSG00000087074.7
PSMD12	DEG	112	117	228	0.68	11.0	ENSG00000197170.5
RND3	DEG	47	78	125	0.69	9.9	ENSG00000115963.9
SBDS	DEG	111	83	194	0.43	8.0	ENSG00000126524.5
SDC4	DEG	89	105	194	1.60	26.0	ENSG00000124145.5
SERPINB9	DEG	76	196	273	2.30	24.0	ENSG00000170542.5
SLC2A1	DEG	109	18	127	0.46	8.5	ENSG00000117394.15
SPRY2	DEG	92	45	137	1.00	16.0	ENSG00000136158.6
TFB2M	DEG	111	10	121	0.55	7.9	ENSG00000162851.6
TFPI2	DEG	111	202	313	5.50	74.0	ENSG00000105825.7
TNFAIP3	DEG	112	72	185	0.81	8.3	ENSG00000118503.10
TNFRSF12A	DEG	46	183	229	2.90	47.0	ENSG00000006327.9
TXNL4B	DEG	120	209	329	0.73	11.0	ENSG00000140830.4
UBALD2	DEG	49	21	70	0.78	12.0	ENSG00000185262.7
UTP3	DEG	112	10	122	0.63	7.6	ENSG00000132467.2
ZCCHC12	DEG	96	64	160	2.70	40.0	ENSG00000174460.3
ZFP36	DEG	49	70	120	2.60	32.0	ENSG00000128016.4
ZNF26	DEG	112	47	159	0.77	9.4	ENSG00000198393.3
ZSWIM6	DEG	114	72	186	0.56	11.0	ENSG00000130449.5
AKIRIN2	ILG	0	2880	2880	0.76	14.0	ENSG00000135334.8
ARC	ILG	0	127	127	2.20	40.0	ENSG00000198576.2
BAIAP2	ILG	0	725	725	1.20	16.0	ENSG00000175866.11
BCL10	ILG	0	142	142	0.98	13.0	ENSG00000142867.8
CLU	ILG	0	457	457	1.60	24.0	ENSG00000120885.15
CMC2	ILG	0	242	242	0.72	14.0	ENSG00000103121.4
CSRNP1	ILG	0	136	136	1.40	15.0	ENSG00000144655.10
CTGF	ILG	0	287	287	1.50	19.0	ENSG00000118523.5
DUSP6	ILG	0	124	124	1.50	17.0	ENSG00000139318.7

Continued on next page.

Table B.3 Gene-wise parameter estimates for IEGs, DEGs, ILGS and SRGs.

Gene	Cluster	Δt	t_{1_2}	r	Log_2 fc	z	ENSEMBLE
EGR4	ILG	0	303	303	1.60	20.0	ENSG00000135625.6
F2RL1	ILG	0	196	196	0.95	14.0	ENSG00000164251.4
FOSL1	ILG	0	139	139	1.00	14.0	ENSG00000175592.4
FOSL2	ILG	0	204	204	0.69	8.3	ENSG00000075426.7
GPR3	ILG	0	129	129	1.40	20.0	ENSG00000181773.6
LY6K	ILG	0	188	188	1.50	16.0	ENSG00000160886.9
MCL1	ILG	0	138	138	0.69	11.0	ENSG00000143384.8
MFSD2A	ILG	0	154	154	1.40	19.0	ENSG00000168389.13
MIDN	ILG	0	186	186	0.90	13.0	ENSG00000167470.8
NPPC	ILG	0	133	133	0.66	9.8	ENSG00000163273.3
NR4A1	ILG	0	2232	2232	2.70	39.0	ENSG00000123358.15
NXF1	ILG	0	226	226	0.47	8.5	ENSG00000162231.9
PNP	ILG	0	273	273	0.46	8.8	ENSG00000198805.7
PTGS2	ILG	0	306	306	2.80	28.0	ENSG00000073756.7
QSOX1	ILG	0	561	561	1.50	18.0	ENSG00000116260.12
SPRY4	ILG	0	131	131	1.70	19.0	ENSG00000187678.8
SYAP1	ILG	0	266	266	0.55	10.0	ENSG00000169895.5
ZYX	ILG	0	225	225	1.00	17.0	ENSG00000159840.11
ABCB1	SRG	114	2880	2994	1.20	17.0	ENSG00000085563.10
AGPAT5	SRG	94	163	257	0.57	8.3	ENSG00000155189.7
ARID3B	SRG	236	46	282	0.82	11.0	ENSG00000179361.13
BZW1	SRG	169	52	221	0.52	9.5	ENSG00000082153.13
CA8	SRG	155	54	210	0.66	9.7	ENSG00000178538.5
CACNA2D1	SRG	99	429	527	0.45	6.7	ENSG00000153956.11
CCNE2	SRG	128	110	238	0.44	6.9	ENSG00000175305.12
CD44	SRG	168	120	288	1.10	15.0	ENSG00000026508.12
CD55	SRG	93	111	204	0.64	12.0	ENSG00000196352.9
CDK7	SRG	113	67	180	0.63	7.8	ENSG00000134058.6
CHAC1	SRG	233	12	244	0.83	12.0	ENSG00000128965.7
CHRM4	SRG	0	343	343	0.91	11.0	ENSG00000180720.6
CHST15	SRG	120	2880	3000	0.82	10.0	ENSG00000182022.13
CLDND1	SRG	155	217	372	0.36	6.8	ENSG00000080822.12
CPEB2	SRG	145	65	210	0.57	6.2	ENSG00000137449.11

Continued on next page.

Table B.3 Gene-wise parameter estimates for IEGs, DEGs, ILGS and SRGs.

Gene	Cluster	Δt	t_{1_2}	r	Log_2 fc	z	ENSEMBLE
CRABP2	SRG	60	253	313	1.10	11.0	ENSG00000143320.4
CYB5R2	SRG	96	1100	1195	1.80	20.0	ENSG00000166394.10
DNAJA1	SRG	0	151	151	0.73	11.0	ENSG00000086061.11
DRAP1	SRG	346	10	356	0.51	9.3	ENSG00000175550.3
ELF4	SRG	150	111	260	1.00	14.0	ENSG00000102034.12
EMP1	SRG	150	140	290	2.50	24.0	ENSG00000134531.5
EN2	SRG	93	78	171	1.40	22.0	ENSG00000164778.4
ENC1	SRG	106	36	142	1.30	15.0	ENSG00000171617.9
EPHA2	SRG	152	122	274	1.60	21.0	ENSG00000142627.9
ERCC1	SRG	100	2880	2980	1.60	19.0	ENSG00000012061.11
FAM3C	SRG	109	80	189	0.64	9.3	ENSG00000196937.6
FBXO32	SRG	104	23	127	1.10	14.0	ENSG00000156804.3
GBE1	SRG	104	1021	1124	1.00	10.0	ENSG00000114480.8
GCH1	SRG	96	113	209	0.69	9.2	ENSG00000131979.14
GNL2	SRG	105	46	151	0.42	7.7	ENSG00000134697.8
GPR137B	SRG	115	112	227	0.70	10.0	ENSG00000077585.9
GPR75	SRG	60	199	259	0.51	6.9	ENSG00000119737.5
HIPK2	SRG	223	38	261	0.39	7.2	ENSG00000064393.11
HK2	SRG	86	78	163	0.45	6.7	ENSG00000159399.5
HMGCS1	SRG	100	210	310	1.20	22.0	ENSG00000112972.10
IFRD1	SRG	137	133	269	0.78	12.0	ENSG00000006652.9
INPP1	SRG	117	142	259	1.10	13.0	ENSG00000151689.8
KCTD13	SRG	159	337	496	0.69	12.0	ENSG00000174943.5
LONRF3	SRG	136	343	479	0.80	11.0	ENSG00000175556.12
LRRC8B	SRG	150	81	231	0.60	8.8	ENSG00000197147.8
LY6G5B	SRG	120	70	190	0.51	8.8	ENSG00000240053.8
NAB2	SRG	44	86	130	1.00	13.0	ENSG00000166886.8
NPAS2	SRG	162	46	208	0.63	9.0	ENSG00000170485.12
NPTX1	SRG	335	16	351	0.76	11.0	ENSG00000171246.5
OTUD4	SRG	235	29	264	0.52	7.5	ENSG00000164164.11
PBDC1	SRG	111	33	143	0.46	5.7	ENSG00000102390.6
PEA15	SRG	168	80	248	0.63	12.0	ENSG00000162734.8
PGBD5	SRG	153	25	178	0.79	9.5	ENSG00000177614.5

Continued on next page.

Table B.3 Gene-wise parameter estimates for IEGs, DEGs, ILGS and SRGs.

Gene	Cluster	Δt	t_{1_2}	r	Log_2 fc	z	ENSEMBLE
PHLDA2	SRG	0	2438	2438	1.20	17.0	ENSG00000181649.5
PNPLA8	SRG	120	155	275	0.54	7.9	ENSG00000135241.12
PRKCD	SRG	164	10	174	0.91	12.0	ENSG00000163932.9
PRNP	SRG	98	90	188	0.37	6.7	ENSG00000171867.12
PXDC1	SRG	98	79	176	1.10	13.0	ENSG00000168994.9
PYGO1	SRG	118	31	149	0.70	9.9	ENSG00000171016.7
RAI2	SRG	109	22	131	1.00	13.0	ENSG00000131831.13
RIOK1	SRG	108	55	163	0.34	5.9	ENSG00000124784.4
RNF19A	SRG	111	21	132	0.37	5.2	ENSG00000034677.7
RNF24	SRG	116	52	168	0.66	9.2	ENSG00000101236.12
ROMO1	SRG	0	118	118	0.81	14.0	ENSG00000125995.11
RPS24	SRG	102	47	148	0.60	11.0	ENSG00000138326.14
SERPINB8	SRG	60	1712	1772	1.80	18.0	ENSG00000166401.9
SERPINE2	SRG	228	46	274	1.20	16.0	ENSG00000135919.8
SERTAD1	SRG	57	17	73	0.76	9.6	ENSG00000197019.4
SGMS2	SRG	146	105	251	1.20	18.0	ENSG00000164023.10
SLC30A7	SRG	106	44	150	0.36	6.8	ENSG00000162695.7
SLC35G2	SRG	108	205	313	0.51	6.7	ENSG00000168917.8
SLC7A1	SRG	95	88	183	0.58	11.0	ENSG00000139514.8
SLC7A5	SRG	0	527	527	0.68	12.0	ENSG00000103257.4
SPOPL	SRG	150	76	225	0.55	6.7	ENSG00000144228.4
SPSB1	SRG	0	296	296	0.92	10.0	ENSG00000171621.9
SRXN1	SRG	60	184	244	0.66	9.5	ENSG00000271303.1
STX11	SRG	93	1085	1178	1.30	16.0	ENSG00000135604.9
TAF13	SRG	111	201	312	1.00	13.0	ENSG00000197780.5
TCF7	SRG	96	2880	2976	0.98	13.0	ENSG00000081059.15
TGFB1	SRG	0	1785	1785	1.40	17.0	ENSG00000105329.5
THBS1	SRG	104	24	128	0.90	7.8	ENSG00000137801.9
TMEM2	SRG	105	100	205	0.81	9.9	ENSG00000135048.9
TMEM88	SRG	3	1446	1449	1.30	14.0	ENSG00000167874.6
TNC	SRG	91	191	282	0.83	11.0	ENSG00000041982.10
TNFRSF10D	SRG	90	130	220	0.66	10.0	ENSG00000173530.5
TOR1B	SRG	87	81	167	0.78	11.0	ENSG00000136816.11

Continued on next page.

Table B.3 Gene-wise parameter estimates for IEGs, DEGs, ILGS and SRGs.

Gene	**Cluster**	**Δt**	**t_{1_2}**	r	**Log_2 fc**	**z**	**ENSEMBLE**
TRIB3	SRG	153	2880	3033	0.82	12.0	ENSG00000101255.6
TRIM9	SRG	112	50	162	0.67	8.5	ENSG00000100505.9
UAP1	SRG	104	69	173	0.76	11.0	ENSG00000117143.9
UBXN8	SRG	148	271	419	0.66	11.0	ENSG00000104691.10
UPP1	SRG	80	2880	2960	1.60	17.0	ENSG00000183696.9
VGF	SRG	289	2880	3169	1.50	18.0	ENSG00000128564.5

End of table.

C Indices

C.1 Bibliography

Abu-Mostafa, Yaser S, Malik Magdon-Ismail, and Hsuan-Tien Lin (Jan. 2012). *Learning from Data: A Short Course*. AMLbook.com.

Adami, Christoph (Apr. 2004). "Information theory in molecular biology". In: *Physics of Life Reviews* 1.1.

Aitken, Stuart et al. (Apr. 2015). "Transcriptional Dynamics Reveal Critical Roles for Non-coding RNAs in the Immediate-Early Response". In: *PLoS Computational Biology* 11.4.

Aksamitiene, Edita, Anatoly B Kiyatkin, and Boris N Kholodenko (Jan. 2012). "Cross-talk between mitogenic Ras/MAPK and survival PI3K/Akt pathways: a fine balance". In: *Biochemical Society Transactions* 40.1.

Alberts, Bruce et al. (Nov. 2014). *Molecular Biology of the Cell, Sixth Edition*. Garland Science.

Alpert, Tara, Lydia Herzel, and Karla M Neugebauer (Mar. 2017). "Perfect timing: splicing and transcription rates in living cells". In: *Wiley Interdisciplinary Reviews: RNA* 8.2.

Amit, Ido et al. (Feb. 2007). "A module of negative feedback regulators defines growth factor signaling". In: *Nature Genetics* 39.4.

Ashburner, Michael et al. (Jan. 1974). "Temporal Control of Puffing Activity in Polytene Chromosomes". In: *Cold Spring Harbor Symposia on Quantitative Biology* 38.0.

Avery, O T, C M Macleod, and M McCarty (Feb. 1944). "Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types: Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated From Pneumococcus Type III". In: *The Journal of experimental medicine* 79.2.

Avraham, R et al. (June 2010). "EGF Decreases the Abundance of MicroRNAs That Restrain Oncogenic Transcription Factors". In: *Science Signaling* 3.124.

Avraham, Roi and Yosef Yarden (Feb. 2011). "Feedback regulation of EGFR signalling: decision making by early and delayed loops". In: *Nature reviews. Molecular cell biology* 12.2.

Baltz, Alexander G et al. (June 2012). "The mRNA-Bound Proteome and Its Global Occupancy Profile on Protein-Coding Transcripts". In: *Molecular cell* 46.5.

Bar-Joseph, Ziv, Anthony Gitter, and Itamar Simon (July 2012). "Studying and modelling dynamic biological processes using time-series gene expression data". In: *Nature Reviews Genetics* 13.8.

Batchelor, Eric et al. (Jan. 2011). "Stimulus-dependent dynamics of p53 in single cells". In: *Molecular Systems Biology* 7.1.

Battich, Nico, Thomas Stoeger, and Lucas Pelkmans (Dec. 2015). "Control of Transcript Variability in Single Mammalian Cells". In: *Cell* 163.7.

Behar, Marcelo and Alexander Hoffmann (Dec. 2010). "Understanding the temporal codes of intra-

cellular signals." In: *Current Opinion in Genetics & Development* 20.6.

Behar, Marcelo et al. (Oct. 2008). "Dose-to-Duration Encoding and Signaling beyond Saturation in Intracellular Signaling Networks". In: *PLoS Computational Biology* 4.10.

Bentler, P M and Douglas G Bonett (Nov. 1980). "Significance tests and goodness of fit in the analysis of covariance structures." In: *Psychological Bulletin* 88.3.

Berlin, C M and R T Schimke (Sept. 1965). "Influence of turnover rates on the responses of enzymes to cortisone." In: *Molecular pharmacology* 1.2.

Bhagwat, Anand S and Christopher R Vakoc (Sept. 2015). "Targeting Transcription Factors in Cancer". In: *Trends in Cancer* 1.1.

Bhalla, Upinder S (Jan. 2003). "Understanding complex signaling networks through models and metaphors." In: *Progress in biophysics and molecular biology* 81.1.

Bhalla, Upinder S, Prahlad T Ram, and Ravi Iyengar (Aug. 2002). "MAP kinase phosphatase as a locus of flexibility in a mitogen-activated protein kinase signaling network." In: *Science* 297.5583.

Bianconi, Eva et al. (Nov. 2013). "An estimation of the number of cells in the human body". In: *Annals of Human Biology* 40.6.

Bister, K and P H Duesberg (Oct. 1979). "Structure and specific sequences of avian erythroblastosis virus RNA: evidence for multiple classes of transforming genes among avian tumor viruses." In: *Proceedings of the National Academy of Sciences of the United States of America* 76.10.

Blackadar, Clarke Brian (2016). "Historical review of the causes of cancer". In: *World Journal of Clinical Oncology* 7.1.

Blume, S W et al. (Nov. 1991). "Mithramycin inhibits SP1 binding and selectively inhibits transcriptional activity of the dihydrofolate reductase gene in vitro and in vivo." In: *Journal of Clinical Investigation* 88.5.

Blüthgen, Nils and Stefan Legewie (2008). "Systems analysis of MAPK signal transduction." In: *Essays in biochemistry* 45.

Bonner, T I, S B Kerby, and P Sutrave (1985). "Structure and biological activity of human homologs of the raf/mil oncogene." In: *Molecular and cellular biology* 5.6.

Box, George E P and Norman Richard Draper (Jan. 1987). *Empirical model-building and response surfaces*. John Wiley & Sons Inc.

Brennan, C M and J A Steitz (Feb. 2001). "HuR and mRNA stability". In: *Cellular and molecular life sciences* 58.2.

Brenner, Sydney (Jan. 2010). "Sequences and consequences". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 365.1537.

Brose, Marcia S et al. (Dec. 2002). "BRAF and RAS Mutations in Human Lung Cancer and Melanoma". In: *Cancer research* 62.23.

Bucheit, Amanda D and Michael A Davies (Nov. 2013). "Emerging insights into resistance to BRAF inhibitors in melanoma." In: *Biochemical pharmacology* 87.3.

Buday, László and Julian Downward (Dec. 2008). "Many faces of Ras activation". In: *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer* 1786.2.

Bugaj, L J et al. (Aug. 2018). "Cancer mutations and targeted drugs can disrupt dynamic signal encoding by the Ras-Erk pathway". In: *Science* 361.6405.

Cagnol, S, E Obberghen-Schilling, and J C Chambard (Mar. 2006). "Prolonged activation of ERK1,2 induces FADD-independent caspase 8 activation and cell death". In: *Apoptosis* 11.3.

Cagnol, Sebastien and Jean-Claude Chambard (Jan. 2010). "ERK and cell death: Mechanisms of ERK-induced cell death – apoptosis, autophagy and senescence". In: *FEBS Journal* 277.1.

Capaldi, Andrew P et al. (Oct. 2008). "Structure and function of a transcriptional network activated by the MAPK Hog1". In: *Nature Genetics* 40.11.

CDC (2018). *U.S. Cancer Statistics Data Visualizations Tool, based on November 2017 submission data (1999-2015)*. www.cdc.gov/cancer/dataviz.

Chapman, Paul B et al. (June 2011). "Improved Survival with Vemurafenib in Melanoma with BRAF V600E Mutation". In: *The New England journal of medicine* 364.26.

Cheng, Christine S et al. (Mar. 2017). “Iterative Modeling Reveals Evidence of Sequential Transcriptional Control Mechanisms.” In: *Cell systems* 4.3.

Chow, Chi-Wing and Roger J Davis (Dec. 2006). “Proteins Kinases: Chromatin-Associated Enzymes?” In: *Cell* 127.5.

Cleary, Michael D et al. (Feb. 2005). “Biosynthetic labeling of RNA with uracil phosphoribosyltransferase allows cell-specific microarray analysis of mRNA synthesis and decay”. In: *Nature biotechnology* 23.2.

Corbalan-Garcia, S et al. (Oct. 1996). “Identification of the mitogen-activated protein kinase phosphorylation sites on human Sos1 that regulate interaction with Grb2.” In: *Molecular and cellular biology* 16.10.

Cox, Adrienne D et al. (Nov. 2014). “Drugging the undruggable RAS: Mission possible?” In: *Nature reviews. Drug discovery* 13.11.

Crick, Francis et al. (1961). *General Nature of the Genetic Code for Proteins.* Macmillan Journals Limited.

Davis, R J (July 1993). “The mitogen-activated protein kinase signal transduction pathway.” In: *The Journal of biological chemistry* 268.20.

Deneke, Carlus, Reinhard Lipowsky, and Angelo Valleriani (Feb. 2013a). “Complex degradation processes lead to non-exponential decay patterns and age-dependent decay rates of messenger RNA”. In: *PloS one* 8.2.

– (Aug. 2013b). “Effect of ribosome shielding on mRNA stability”. In: *Physical Biology* 10.4.

Der, C J, T G Krontiris, and G M Cooper (June 1982). “Transforming genes of human bladder and lung carcinoma cell lines are homologous to the ras genes of Harvey and Kirsten sarcoma viruses.” In: *Proceedings of the National Academy of Sciences of the United States of America* 79.11.

Dhillon, A S et al. (May 2007). “MAP kinase signalling pathways in cancer”. In: *Oncogene* 26.22.

Dhillon, Amardeep S et al. (May 2002). “Cyclic AMP-dependent kinase regulates Raf-1 kinase mainly by phosphorylation of serine 259.” In: *Molecular and cellular biology* 22.10.

Dijkmans, T F et al. (Jan. 2009). “Identification of new Nerve Growth Factor-responsive immediate-early genes”. In: *Brain Research* 1249.

Dobin, Alexander et al. (Jan. 2013). “STAR: ultrafast universal RNA-seq aligner.” In: *Bioinformatics (Oxford, England)* 29.1.

Dodt, Matthias et al. (Dec. 2012). “FLEXBAR-Flexible Barcode and Adapter Processing for Next-Generation Sequencing Platforms.” In: *Biology* 1.3.

Dölken, Lars et al. (Sept. 2008). “High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay.” In: *RNA* 14.9.

Downward, J et al. (Feb. 1984). “Close similarity of epidermal growth factor receptor and v-erb-B oncogene protein sequences.” In: *Nature* 307.5951.

Downward, Julian (Jan. 2003). “Targeting RAS signalling pathways in cancer therapy”. In: *Nature Reviews Cancer* 3.1.

Druker, Brian J (Dec. 2003). “David A. Karnofsky Award lecture. Imatinib as a paradigm of targeted therapies.” In: *Journal of clinical oncology* 21.23.

Druker, Brian J et al. (Dec. 2006). “Five-year follow-up of patients receiving imatinib for chronic myeloid leukemia.” In: *The New England journal of medicine* 355.23.

Duesberg, P H and P K Vogt (Dec. 1970). “Differences between the ribonucleic acids of transforming and nontransforming avian tumor viruses.” In: *Proceedings of the National Academy of Sciences of the United States of America* 67.4.

Elkon, Ran et al. (Apr. 2010). “Major role for mRNA stability in shaping the kinetics of gene induction.” In: *BMC genomics* 11.1.

Elowitz, Michael B et al. (Aug. 2002). “Stochastic Gene Expression in a Single Cell”. In: *Science* 297.5584.

Erhardt, P, E J Schremser, and G M Cooper (Aug. 1999). “B-Raf inhibits programmed cell death downstream of cytochrome c release from mitochondria by activating the MEK/Erk pathway.” In: *Molecular and cellular biology* 19.8.

Escudier, Bernard et al. (Jan. 2007). “Sorafenib in Advanced Clear-Cell Renal-Cell Carcinoma”. In: *The New England journal of medicine* 356.2.

Fabian, M R and N Sonenberg (2010). “Regulation of mRNA translation and stability by microRNAs”. In: *Annual Review of Biochemistry* 79.

Feldman, Morris E and Yosef Yarden (Aug. 2014). “Steering tumor progression through the transcriptional response to growth factors and stroma.” In: *FEBS letters* 588.15.

Ferrell, James E and Eric M Machleder (May 1998). “The Biochemical Basis of an All-or-None Cell Fate Switch in Xenopus Oocytes”. In: *Science* 280.5365.

Fowler, Trent, Ranjan Sen, and Ananda L Roy (Nov. 2011). “Regulation of Primary Response Genes”. In: *Molecular cell* 44.3.

Frevel, Mathias A E et al. (Jan. 2003). “p38 Mitogen-activated protein kinase-dependent and -independent signaling of mRNA stability of AU-rich element-containing transcripts.” In: *Molecular and cellular biology* 23.2.

Friday, Bret B et al. (Aug. 2008). “BRAF V600E disrupts AZD6244-induced abrogation of negative feedback pathways between extracellular signal-regulated kinase and Raf proteins.” In: *Cancer research* 68.15.

Friedel, Caroline C et al. (Sept. 2009). “Conserved principles of mammalian transcriptional regulation revealed by RNA half-life.” In: *Nucleic acids research* 37.17.

Fritsche-Guenther, Raphaela et al. (May 2011). “Strong negative feedback from Erk to Raf confers robustness to MAPK signalling.” In: *Molecular Systems Biology* 7.1.

Fritsche-Guenther, Raphaela et al. (Jan. 2016). “Effects of RAF inhibitors on PI3K/AKT signalling depend on mutational status of the RAS/RAF signalling axis”. In: *Oncotarget* 7.7.

Fujita, Toshitsugu, Isabelle Piuz, and Werner Schlegel (Jan. 2009). “Negative elongation factor NELF controls transcription of immediate early genes in a stimulus-specific manner.” In: *Experimental cell research* 315.2.

Fulda, S and K-M Debatin (Aug. 2006). “Extrinsic versus intrinsic apoptosis pathways in anticancer chemotherapy”. In: *Oncogene* 25.34.

Gao, J et al. (Apr. 2013). “Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal”. In: *Science Signaling* 6.269.

Gillies, Taryn E et al. (Dec. 2017). “Linear Integration of ERK Activity Predominates over Persistence Detection in Fra-1 Regulation”. In: *Cell systems* 5.6.

Goldman, David E (Sept. 1943). “Potential, Impedance, and Rectification in Membranes”. In: *The Journal of general physiology* 27.1.

Gómez-Schiavon, Mariana et al. (2017). “BayFish: Bayesian inference of transcription dynamics from population snapshots of single-molecule RNA FISH in single cells”. In: *Genome biology* 18.1.

Gorini, Luigi and Werner K Maas (Jan. 1957). “The potential for the formation of a biosynthetic enzyme in Escherichia coli”. In: *Biochimica et biophysica acta* 25.

Gu, Hua, Yong-Rui Zou, and Klaus Rajewsky (June 1993). “Independent control of immunoglobulin switch recombination at individual switch regions evidenced through Cre-loxP-mediated gene targeting”. In: *Cell* 73.6.

Hanahan, Douglas and Robert A Weinberg (Mar. 2011). “Hallmarks of Cancer: The Next Generation”. In: *Cell* 144.5.

Hannon, Gregory J (July 2002). “RNA interference”. In: *Nature* 418.6894.

Hao, Nan and Erin K O’Shea (Jan. 2012). “Signal-dependent dynamics of transcription factor translocation controls gene expression”. In: *Nature structural and molecular biology* 19.1.

Hao, S and D Baltimore (July 2013). “RNA splicing regulates the temporal order of TNF-induced gene expression”. In: *Proceedings of the National Academy of Sciences of the United States of America* 110.29.

Hao, Shengli and David Baltimore (Mar. 2009). “The stability of mRNA influences the temporal order of the induction of genes encoding inflammatory molecules.” In: *Nature immunology* 10.3.

Hargreaves, Diana C, Tiffany Horng, and Ruslan Medzhitov (July 2009). “Control of inducible gene expression by signal-dependent transcriptional elongation.” In: *Cell* 138.1.

Hargrove, James L, Martin G Hulsey, and Elmus G Beale (Dec. 1991). "The kinetics of mammalian gene expression". In: *BioEssays : news and reviews in molecular, cellular and developmental biology* 13.12.

Hauschild, Axel et al. (July 2012). "Dabrafenib in BRAF-mutated metastatic melanoma: a multicentre, open-label, phase 3 randomised controlled trial". In: *The Lancet* 380.9839.

Healy, Shannon, Protiti Khan, and James R Davie (Jan. 2013). "Immediate early response genes and cell transformation". In: *Pharmacology & Therapeutics* 137.1.

Hershey, A D and M Chase (May 1952). "Independent functions of viral protein and nucleic acid in growth of bacteriophage." In: *The Journal of general physiology* 36.1.

Herzog, Veronika A et al. (Dec. 2017). "Thiol-linked alkylation of RNA to assess expression dynamics." In: *Nature methods* 14.12.

Hoagland, Mahlon B et al. (1958). "A soluble ribonucleic acid intermediate in protein synthesis". In: *Journal of Biological Chemistry* 231.1.

Hodgkin, A L and A F Huxley (Aug. 1952). "A quantitative description of membrane current and its application to conduction and excitation in nerve." In: *The Journal of Physiology* 117.4.

Hoffmann, Alexander et al. (Nov. 2002). "The IκB-NF-κB Signaling Module: Temporal Control and Selective Gene Activation". In: *Science* 298.5596.

Hsu, Patrick D, Eric S Lander, and Feng Zhang (June 2014). "Development and Applications of CRISPR-Cas9 for Genome Engineering". In: *Cell* 157.6.

Hu, Shaohui et al. (Oct. 2009). "Profiling the Human Protein-DNA Interactome Reveals ERK2 as a Transcriptional Repressor of Interferon Signaling". In: *Cell* 139.3.

Huang, C Y and J E Ferrell (Sept. 1996). "Ultrasensitivity in the mitogen-activated protein kinase cascade." In: *Proceedings of the National Academy of Sciences of the United States of America* 93.19.

Irizarry, R A (Apr. 2003). "Exploration, normalization, and summaries of high density oligonucleotide array probe level data". In: *Biostatistics* 4.2.

Jacob, François and Jacques Monod (June 1961). "Genetic regulatory mechanisms in the synthesis of proteins". In: *Journal of Molecular Biology* 3.3.

Jinek, Martin et al. (Aug. 2012). "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity." In: *Science* 337.6096.

Johnson, Heath E, Stanislav Y Shvartsman, and Jared E Toettcher (June 2018). "Signaling dynamics control cell fate in the early Drosophila embryo". In: *bioRxiv*.

Jürges, Christopher, Lars Dölken, and Florian Erhard (July 2018). "Dissecting newly transcribed and old RNA using GRAND-SLAM". In: *Bioinformatics (Oxford, England)* 34.13.

Kassambara, A, M Kosinski, and P Biecek (2017). "survminer: Drawing Survival Curves using'ggplot2'". In: *CRAN* 1.

Kholodenko, Boris N (Mar. 2006). "Cell-signalling dynamics in time and space". In: *Nature reviews. Molecular cell biology* 7.3.

Kholodenko, Boris N, John F Hancock, and Walter Kolch (June 2010). "Signalling ballet in space and time". In: *Nature reviews. Molecular cell biology* 11.6.

Kidger, Andrew M and Stephen M Keyse (Feb. 2016). "The regulation of oncogenic Ras/ERK signalling by dual-specificity mitogen activated protein kinase phosphatases (MKPs)". In: *Seminars in Cell & Developmental Biology* 50.

Klinger, Bertram et al. (2013). "Network quantification of EGFR signaling unveils potential for targeted combination therapy." In: *Molecular Systems Biology* 9.

Kolch, Walter et al. (Sept. 2015). "The dynamic control of signal transduction networks in cancer cells." In: *Nature Reviews Cancer* 15.9.

Köster, Johannes and Sven Rahmann (Oct. 2012). "Snakemake - a scalable bioinformatics workflow engine." In: *Bioinformatics (Oxford, England)* 28.19.

Kramer, G and H Hilz (June 1971). "A new approach to the determination of mRNA half-life time in HeLa cells without the use of actinomycin." In: *Hoppe-Seyler's Zeitschrift fur physiologische Chemie* 352.6.

Kreutz, Clemens and Jens Timmer (Feb. 2009). "Systems biology: experimental design". In: *FEBS Journal* 276.4.

Lake, David, Sonia A L Corrêa, and Jürgen Müller (2016). "Negative feedback regulation of the ERK1/2 MAPK pathway". In: *Cellular and Molecular Life Sciences* 73.23.

Larkin, James et al. (Nov. 2014). "Combined Vemurafenib and Cobimetinib in BRAF-Mutated Melanoma". In: *The New England journal of medicine* 371.20.

Lavoie, Hugo and Marc Therrien (May 2015). "Regulation of RAF protein kinases in ERK signalling". In: *Nature reviews. Molecular cell biology* 16.5.

Lehmann, K et al. (Oct. 2000). "Raf induces TGF-beta production while blocking its apoptotic but not invasive responses: a mechanism leading to increased malignancy in epithelial cells." In: *Genes & Development* 14.20.

Lenormand, P et al. (Aug. 1998). "Growth factor-induced p42/p44 MAPK nuclear translocation and retention requires both MAPK activation and neosynthesis of nuclear anchoring proteins." In: *The Journal of cell biology* 142.3.

Levinthal, Cyrus, Alex Keynan, and Akiko Higa (Sept. 1962). "Messenger RNA Turnover and Protein Synthesis in B. subtilis Inhibited by Actinomycin D". In: *Proceedings of the National Academy of Sciences of the United States of America* 48.9.

Li, Bo and Colin N Dewey (Dec. 2011). "RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome". In: *BMC bioinformatics* 12.1.

Liao, Yang, Gordon K Smyth, and Wei Shi (Apr. 2014). "featureCounts: an efficient general purpose program for assigning sequence reads to genomic features." In: *Bioinformatics (Oxford, England)* 30.7.

Lito, Piro et al. (Feb. 2016). "Allele-specific inhibitors inactivate mutant KRAS G12C by a trapping mechanism". In: *Science* 351.6273.

Litvak, Vladimir et al. (Apr. 2009). "Function of C/EBPδ in a regulatory circuit that discriminates between transient and persistent TLR4-induced signals". In: *Nature immunology* 10.4.

Llovet, Josep M et al. (July 2008). "Sorafenib in Advanced Hepatocellular Carcinoma". In: *The New England journal of medicine* 359.4.

Loewer, Alexander et al. (July 2010). "Basal Dynamics of p53 Reveal Transcriptionally Attenuated Pulses in Cycling Cells". In: *Cell* 142.1.

Lynch, Thomas J et al. (May 2004). "Activating Mutations in the Epidermal Growth Factor Receptor Underlying Responsiveness of Non–Small-Cell Lung Cancer to Gefitinib". In: *The New England journal of medicine* 350.21.

Mann, Michael J (Nov. 2005). "Transcription factor decoys: a new model for disease intervention." In: *Annals of the New York Academy of Sciences* 1058.

Mariathasan, Sanjeev et al. (Nov. 2001). "Duration and Strength of Extracellular Signal-Regulated Kinase Signals Are Altered During Positive Versus Negative Thymocyte Selection". In: *The Journal of Immunology* 167.9.

Marshall, C J (Jan. 1995). "Specificity of receptor tyrosine kinase signaling: Transient versus sustained extracellular signal-regulated kinase activation". In: *Cell* 80.2.

Martin, Patrick and Philippe Pognonec (Jan. 2010). "ERK and cell death: cadmium toxicity, sustained ERK activation and cell death". In: *FEBS Journal* 277.1.

Marzi, Matteo J et al. (Jan. 2016). "Degradation dynamics of microRNAs revealed by a novel pulse-chase approach". In: *Genome Research*.

Mason, Jacqueline M et al. (Jan. 2006). "Sprouty proteins: multifaceted negative-feedback regulators of receptor tyrosine kinase signaling". In: *Trends in Cell Biology* 16.1.

Matallanas, D et al. (July 2011). "Raf Family Kinases: Old Dogs Have Learned New Tricks". In: *Genes & Cancer* 2.3.

Maughan, Tim (May 2017). "The Promise and the Hype of 'Personalised Medicine'". In: *The New Bioethics* 23.1.

Mauro, M J and B J Druker (2001). "STI571: targeting BCR-ABL as therapy for CML." In: *The Oncologist* 6.3.

Mazzarello, P (May 1999). "A unifying concept: the history of cell theory". In: *Nature Cell Biology* 1.1.

McMahon, M (2001). “Steroid receptor fusion proteins for conditional activation of Raf-MEK-ERK signaling pathway”. In: *Regulators and Effectors of Small Gtpases, Pt F, Ras Family I* 332.

Melvin, W T et al. (Dec. 1978). “Incorporation of 6-thioguanosine and 4-thiouridine into RNA. Application to isolation of newly synthesised RNA by affinity chromatography.” In: *European journal of biochemistry* 92.2.

Mendoza, Michelle C, E Emrah Er, and John Blenis (June 2011). “The Ras-ERK and PI3K-mTOR pathways: cross-talk and compensation”. In: *Trends in biochemical sciences* 36.6.

Menten, Leonor and M I Michaelis (1913). “Die kinetik der invertinwirkung”. In: *Biochem Z* 49.

Miller, Christian et al. (Jan. 2011). “Dynamic transcriptome analysis measures rates of mRNA synthesis and decay in yeast”. In: *Molecular Systems Biology* 7.1.

Misale, S, S Arena, and S Lamba (2014). “Blockade of EGFR and MEK Intercepts Heterogeneous Mechanisms of Acquired Resistance to Anti-EGFR Therapies in Colorectal Cancer”. In: *Science Translational Medicine* 6.224.

Monod, J, A M Pappenheimer Jr., and G Cohen-Bazire (Jan. 1952). “La cinétique de la biosynthèse de la β-galactosidase chez E. coli considérée comme fonction de la croissance”. In: *Biochimica et biophysica acta* 9.

Moore, Melissa J and Nick J Proudfoot (Feb. 2009). “Pre-mRNA Processing Reaches Back to Transcription and Ahead to Translation”. In: *Cell* 136.4.

Mukherjee, Neelanjan et al. (Jan. 2014). “Global target mRNA specification and regulation by the RNA-binding protein ZFP36”. In: *Genome biology* 15.1.

Munsky, Brian, Gregor Neuert, and Alexander van Oudenaarden (Apr. 2012). “Using Gene Expression Noise to Understand Gene Regulation”. In: *Science* 336.6078.

Murphy, L O, J P MacKeigan, and J Blenis (Jan. 2004). “A network of immediate early gene products propagates subtle differences in mitogen-activated protein kinase signal amplitude and duration”. In: *Molecular and cellular biology* 24.1.

Murphy, Leon O and John Blenis (May 2006). “MAPK signal specificity: the right place at the right time”. In: *Trends in biochemical sciences* 31.5.

Murphy, Leon O et al. (Aug. 2002). “Molecular interpretation of ERK signal duration by immediate early gene products”. In: *Nature Cell Biology* 4.8.

Nagashima, T et al. (Nov. 2006). “Quantitative Transcriptional Control of ErbB Receptor Signaling Undergoes Graded to Biphasic Response for Cell Differentiation”. In: *Journal of Biological Chemistry* 282.6.

Nagashima, Takeshi et al. (Aug. 2009). “Mutation of epidermal growth factor receptor is associated with MIG6 expression”. In: *FEBS Journal* 276.18.

Nagashima, Takeshi et al. (Feb. 2015). “Feedforward regulation of mRNA stability by prolonged extracellular signal-regulated kinase activity.” In: *FEBS Journal* 282.4.

Neymotin, Benjamin, Rodoniki Athanasiadou, and David Gresham (Oct. 2014). “Determination of in vivo RNA kinetics using RATE-seq.” In: *RNA* 20.10.

Northwood, I C et al. (Aug. 1991). “Isolation and characterization of two growth factor-stimulated protein kinases that phosphorylate the epidermal growth factor receptor at threonine 669.” In: *The Journal of biological chemistry* 266.23.

Oda, Kanae et al. (May 2005). “A comprehensive pathway map of epidermal growth factor receptor signaling”. In: *Molecular Systems Biology* 1.1.

Offermann, Barbara et al. (Apr. 2016). “Boolean Modeling Reveals the Necessity of Transcriptional Regulation for Bistability in PC12 Cell Differentiation”. In: *Frontiers in Genetics* 7.355.

Pan, J G and T W Mak (Apr. 2007). “Metabolic Targeting as an Anticancer Strategy: Dawn of a New Era?” In: *Science Signaling* 2007.381.

Pandolfi, P P (May 2001). “Transcription therapy for cancer.” In: *Oncogene* 20.24.

Parikh, Jignesh R et al. (July 2010). “Discovering causal signaling pathways through gene-expression patterns.” In: *Nucleic acids research* 38.Web Server issue.

Pauly, P J (1987). *Controlling life: Jacques Loeb and the engineering ideal in biology.*

Pellegrino, Michael J and Philip J S Stork (Dec. 2006). "Sustained activation of extracellular signal-regulated kinase by nerve growth factor regulates c-fos protein stabilization and transactivation in PC12 cells". In: *Journal of Neurochemistry* 99.6.

Penman, Sheldon et al. (May 1963). "Polyribosomes in Normal and Poliovirus-Infected HeLa Cells and their Relationship to Messenger-RNA". In: *Proceedings of the National Academy of Sciences of the United States of America* 49.5.

Petersen, N S, C S McLaughlin, and D P Nierlich (Mar. 1976). "Half life of yeast messenger RNA". In: *Nature* 260.5546.

Pokholok, D K (July 2006). "Activated Signal Transduction Kinases Frequently Occupy Target Genes". In: *Science* 313.5786.

Porebska, I, A Harlozińska, and T Bojarowski (Mar. 2000). "Expression of the tyrosine kinase activity growth factor receptors (EGFR, ERB B2, ERB B3) in colorectal adenocarcinomas and adenomas." In: *Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine* 21.2.

Porter, Joshua R, Brian E Fisher, and Eric Batchelor (Apr. 2016). "p53 Pulses Diversify Target Gene Expression Dynamics in an mRNA Half-Life-Dependent Manner and Delineate Co-regulated Target Gene Subnetworks." In: *Cell systems* 2.4.

Posas, F (Mar. 2000). "The Transcriptional Response of Yeast to Saline Stress". In: *Journal of Biological Chemistry* 275.23.

Pretis, Stefano de et al. (May 2015). "INSPEcT: a Computational Tool to Infer mRNA Synthesis, Processing and Degradation Dynamics from RNA- and 4sU-seq Time Course Experiments". In: *Bioinformatics (Oxford, England)* 31.17.

Price, V E et al. (Nov. 1962). "The kinetics of catalase synthesis and destruction in vivo." In: *The Journal of biological chemistry* 237.11.

Purvis, Jeremy E and Galit Lahav (Feb. 2013). "Encoding and Decoding Cellular Information through Signaling Dynamics". In: *Cell* 152.5.

Rabani, Michal et al. (May 2011). "Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells." In: *Nature biotechnology* 29.5.

Rabani, Michal et al. (Dec. 2014). "High-Resolution Sequencing and Modeling Identifies Distinct Dynamic RNA Regulatory Strategies". In: *Cell* 159.7.

Raghavan, Arvind and Paul R Bohjanen (Aug. 2004). "Microarray-based analyses of mRNA decay in the regulation of mammalian gene expression." In: *Briefings in functional genomics & proteomics* 3.2.

Raghavan, Arvind et al. (Dec. 2002). "Genome-wide analysis of mRNA decay in resting and activated primary human T lymphocytes." In: *Nucleic acids research* 30.24.

Rajalingam, Krishnaraj et al. (Aug. 2007). "Ras oncogenes and their downstream targets". In: *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* 1773.8.

Ramaswamy, S et al. (Dec. 2001). "Multiclass cancer diagnosis using tumor gene expression signatures." In: *Proceedings of the National Academy of Sciences of the United States of America* 98.26.

Ramirez-Carrozzi, Vladimir R et al. (July 2009). "A unifying model for the selective regulation of inducible transcription by CpG islands and nucleosome remodeling." In: *Cell* 138.1.

Rapp, U R and M D Goldsborough (1983). "Structure and biological activity of v-raf, a unique oncogene transduced by a retrovirus". In: *Proceedings of the National Academy of Sciences USA* 80.

Reynolds, Andrew S (June 2018). *The Third Lens. Metaphor and the Creation of Modern Cell Biology.* University of Chicago Press.

Rhee, Alex, Raymond Cheong, and Andre Levchenko (Aug. 2012). "The application of information theory to biochemical signaling systems". In: *Physical Biology* 9.4.

Rizzi, Federica et al. (May 2009). "Clusterin is a short half-life, poly-ubiquitinated protein, which controls the fate of prostate cancer cells." In: *Journal of cellular physiology* 219.2.

Robert, Caroline et al. (Jan. 2015). "Improved overall survival in melanoma with combined dabrafenib and trametinib." In: *The New England journal of medicine* 372.1.

Roberts, P J and C J Der (May 2007). "Targeting the Raf-MEK-ERK mitogen-activated protein kinase cascade for the treatment of cancer". In: *Oncogene* 26.22.

Robinson, Mark D, Davis J McCarthy, and Gordon K Smyth (Jan. 2010). "edgeR: a Bioconductor package for differential expression analysis of digital gene expression data". In: *Bioinformatics (Oxford, England)* 26.1.

Roeder, Robert G (Oct. 2003). "The eukaryotic transcriptional machinery: complexities and mechanisms unforeseen". In: *Nature Medicine* 9.10.

Ryu, Hyunryul et al. (Nov. 2015). "Frequency modulation of ERK activation dynamics rewires cell fate". In: *Molecular Systems Biology* 11.11.

Saeki, Yuko et al. (2009). "Ligand-specific sequential regulation of transcription factors for differentiation of MCF-7 cells." In: *BMC genomics* 10.1.

Saez, E et al. (Oct. 1997). "Inducible gene expression in mammalian cells and transgenic mice." In: *Current opinion in biotechnology* 8.5.

Samatar, Ahmed A and Poulikos I Poulikakos (Dec. 2014). "Targeting RAS-ERK signalling in cancer: promises and challenges." In: *Nature reviews. Drug discovery* 13.12.

Samuels, M L et al. (Oct. 1993). "Conditional transformation of cells and rapid activation of the mitogen-activated protein kinase cascade by an estradiol-dependent human raf-1 protein kinase." In: *Molecular and cellular biology* 13.10.

Santos, Silvia D M, Peter J Verveer, and Philippe I H Bastiaens (Mar. 2007). "Growth factor-induced MAPK network topology shapes Erk response determining PC-12 cell fate". In: *Nature Cell Biology* 9.3.

Sasagawa, Satoru et al. (Apr. 2005). "Prediction and validation of the distinct dynamics of transient and sustained ERK activation". In: *Nature Cell Biology* 7.4.

Sato, Kanae et al. (Aug. 2013). "23822636". In: *Cancer Science* 104.10.

Schofield, Jeremy A et al. (Jan. 2018). "TimeLapse-seq: adding a temporal dimension to RNA sequencing through nucleoside recoding". In: *Nature methods* 14.

Schueler, Markus et al. (Jan. 2014). "Differential protein occupancy profiling of the mRNA transcriptome". In: *Genome biology* 15.1.

Schulze, A et al. (Apr. 2001). "Analysis of the transcriptional program induced by Raf in epithelial cells." In: *Genes & Development* 15.8.

Schwalb, Björn et al. (Mar. 2012). "Measurement of genome-wide RNA synthesis and decay rates with Dynamic Transcriptome Analysis (DTA)." In: *Bioinformatics (Oxford, England)* 28.6.

Schwalb, Björn et al. (June 2016). "TT-seq maps the human transient transcriptome". In: *Science* 352.6290.

Schwanhäusser, Björn et al. (May 2011). "Global quantification of mammalian gene expression control." In: *Nature* 473.7347.

Senecal, Adrien et al. (Oct. 2014). "Transcription Factors Modulate c-Fos Transcriptional Bursts". In: *Cell Reports* 8.1.

Shain, A Hunter and Boris C Bastian (Apr. 2016). "From melanocytes to melanomas". In: *Nature Publishing Group*.

Shalem, Ophir et al. (2008). "Transient transcriptional responses to stress are generated by opposing effects of mRNA production and degradation." In: *Molecular Systems Biology* 4.

Shannon, C E (1948). *A Mathematical Theory of Communication*. Bell System Technical Journal.

Sharp, L L et al. (Nov. 1997). "The influence of the MAPK pathway on T cell lineage commitment." In: *Immunity* 7.5.

Shaw, Gerry et al. (June 2002). "Preferential transformation of human neuronal cells by human adenoviruses and the origin of HEK 293 cells." In: *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* 16.8.

Sheiness, D, L Puckett, and J E Darnell (Mar. 1975). "Possible relationship of poly(A) shortening to mRNA turnover". In: *Proceedings of the National Academy of Sciences of the United States of America* 72.3.

Shih, Thomas Y et al. (July 1979). "Identification of a sarcoma virus-coded phosphoprotein in non-

producer cells transformed by Kirsten or Harvey murine sarcoma virus". In: *Virology* 96.1.

Shiraishi, Y, S Kimura, and M Okada (Apr. 2010). "Inferring cluster-based networks from differently stimulated multiple time-course gene expression data". In: *Bioinformatics (Oxford, England)* 26.8.

Shiratsuchi, Toru, Hiroaki Ishibashi, and Kanemitsu Shirasuna (Dec. 2002). "Inhibition of epidermal growth factor-induced invasion by dexamethasone and AP-1 decoy in human squamous cell carcinoma cell lines." In: *Journal of cellular physiology* 193.3.

Singer, R H and S Penman (Aug. 1973). "Messenger RNA in HeLa cells: kinetics of formation and decay." In: *Journal of Molecular Biology* 78.2.

Smolewski, Piotr (Mar. 2008). "Terameprocol, a novel site-specific transcription inhibitor with anticancer activity." In: *IDrugs : the investigational drugs journal* 11.3.

Snider, Janice (June 2014). "Maternal vitamin D levels and early childhood caries may be linked, researchers find." In: *Journal of the American Dental Association (1939)* 145.6.

Stanciu, Madalina and Donald B DeFranco (Feb. 2002). "Prolonged nuclear retention of activated extracellular signal-regulated protein kinase promotes cell death generated by oxidative toxicity or proteasome inhibition in a neuronal cell line." In: *The Journal of biological chemistry* 277.6.

Stehelin, D et al. (Mar. 1976). "DNA related to the transforming gene(s) of avian sarcoma viruses is present in normal avian DNA". In: *Nature* 260.5547.

Stelniec-Klotz, Iwona et al. (2012). "Reverse engineering a hierarchical regulatory network downstream of oncogenic KRAS." In: *Molecular Systems Biology* 8.1.

Sternberg, Nat and Daniel Hamilton (Aug. 1981). "Bacteriophage P1 site-specific recombination: I. Recombination between loxP sites". In: *Journal of Molecular Biology* 150.4.

Subramaniam, Srinivasa and Klaus Unsicker (Jan. 2010). "ERK and cell death: ERK1/2 in neuronal death." In: *FEBS Journal* 277.1.

Sugiura, Reiko et al. (2011). "Role of RNA-Binding Proteins in MAPK Signal Transduction Pathway." In: *Journal of signal transduction* 2011.12.

Sun, Mai et al. (July 2012). "Comparative dynamic transcriptome analysis (cDTA) reveals mutual feedback between mRNA synthesis and degradation." In: *Genome Research* 22.7.

Suter, David M et al. (Apr. 2011). "Mammalian Genes Are Transcribed with Widely Different Bursting Kinetics". In: *Science* 332.6028.

Tani, Hidenori and Nobuyoshi Akimitsu (Oct. 2014). "Genome-wide technology for determining RNA stability in mammalian cells". In: *RNA biology* 9.10.

Tarutani, M et al. (2003). "Inducible activation of Ras and Raf in adult epidermis". In: *Cancer research*.

Teixeiro, Emma and Mark A Daniels (Jan. 2010). "ERK and cell death: ERK location and T cell selection." In: *FEBS Journal* 277.1.

Thiel, Gerald, Myriam Ekici, and Oliver G Rössler (Apr. 2009). "Regulation of cellular proliferation, differentiation and cell death by activated Raf". In: *Cell Communication and Signaling* 7.1.

Tischer, Doug and Orion D Weiner (Aug. 2014). "Illuminating cell signalling with optogenetic tools". In: *Nature reviews. Molecular cell biology* 15.8.

Toettcher, Jared E, Orion D Weiner, and Wendell A Lim (Dec. 2013). "Using Optogenetics to Interrogate the Dynamic Control of Signal Transmission by the Ras/Erk Module". In: *Cell* 155.6.

Traverse, S et al. (Dec. 1992). "Sustained activation of the mitogen-activated protein (MAP) kinase cascade may be required for differentiation of PC12 cells. Comparison of the effects of nerve growth factor and epidermal growth factor." In: *The Biochemical journal* 288 (Pt 2).

Tsao, Ming-Sound et al. (July 2005). "Erlotinib in Lung Cancer — Molecular and Clinical Predictors of Outcome". In: *The New England journal of medicine* 353.2.

Tullai, J W et al. (Aug. 2007). "Immediate-Early and Delayed Primary Response Genes Are Distinct in Function and Genomic Architecture". In: *Journal of Biological Chemistry* 282.33.

Tullai, John W et al. (May 2004). “Identification of transcription factor binding sites upstream of human genes regulated by the phosphatidylinositol 3-kinase and MEK/ERK signaling pathways.” In: *The Journal of biological chemistry* 279.19.

Ueki, K et al. (June 1994). “Feedback regulation of mitogen-activated protein kinase kinase kinase activity of c-Raf-1 by insulin and phorbol ester stimulation.” In: *The Journal of biological chemistry* 269.22.

Uhlitz, Florian et al. (May 2017). “An immediate-late gene expression module decodes ERK signal duration.” In: *Molecular Systems Biology* 13.5.

Ünal, Evrim B, Florian Uhlitz, and Nils Blüthgen (July 2017). “A compendium of ERK targets”. In: *FEBS letters* 60.

Vanneman, Matthew and Glenn Dranoff (Apr. 2012). “Combining immunotherapy and targeted therapies in cancer treatment”. In: *Nature Reviews Cancer* 12.4.

Veer, Laura J van ’t et al. (Jan. 2002). “Gene expression profiling predicts clinical outcome of breast cancer”. In: *Nature* 415.6871.

Villaseñor, Roberto et al. (Feb. 2015). “Regulation of EGFR signal transduction by analogue-to-digital conversion in endosomes”. In: *eLife* 4.

Virchow, R (1859). *Die Cellularpathologie in ihrer Begründung auf physiologische und pathologische Gewebelehre*. Vorlesungen über Pathologie. Hirschwald.

Vogelstein, Bert et al. (Mar. 2013). “Cancer genome landscapes.” In: *Science* 339.6127.

Vogt, Peter K (Sept. 2012). “Retroviral oncogenes: a historical primer”. In: *Nature Reviews Cancer* 12.9.

Volterra, Vito (Oct. 1926). “Fluctuations in the Abundance of a Species considered Mathematically”. In: *Nature* 118.2972.

Wada, Takeo and Attila Becskei (Dec. 2017). “Impact of Methods on the Measurement of mRNA Turnover.” In: *International journal of molecular sciences* 18.12.

Wagle, Nikhil et al. (Aug. 2011). “Dissecting Therapeutic Resistance to RAF Inhibition in Melanoma by Tumor Genomic Profiling”. In: *Journal of Clinical Oncology* 29.22.

Walboomers, J M et al. (Sept. 1999). “Human papillomavirus is a necessary cause of invasive cervical cancer worldwide.” In: *The Journal of pathology* 189.1.

Waltermann, Christian and Edda Klipp (Oct. 2011). “Information theory based approaches to cellular signaling.” In: *Biochimica et biophysica acta* 1810.10.

Wang, X, J L Martindale, and N J Holbrook (Dec. 2000). “Requirement for ERK activation in cisplatin-induced apoptosis.” In: *The Journal of biological chemistry* 275.50.

Wang, Zhibin et al. (Sept. 2009). “Genome-wide Mapping of HATs and HDACs Reveals Distinct Functions in Active and Inactive Genes”. In: *Cell* 138.5.

Welsch, Matthew E et al. (Feb. 2017). “Multivalent Small-Molecule Pan-RAS Inhibitors”. In: *Cell* 168.5.

Wend, Sabrina et al. (May 2014). “Optogenetic Control of Protein Kinase Activity in Mammalian Cells”. In: *ACS Synthetic Biology* 3.5.

West, Alison C and Ricky W Johnstone (Jan. 2014). “New and emerging HDAC inhibitors for cancer treatment.” In: *The Journal of Clinical Investigation* 124.1.

Whalen, A M et al. (Apr. 1997). “Megakaryocytic differentiation induced by constitutive activation of mitogen-activated protein kinase kinase.” In: *Molecular and cellular biology* 17.4.

Whitmarsh, Alan J (Aug. 2007). “Regulation of gene transcription by mitogen-activated protein kinase signaling pathways”. In: *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* 1773.8.

– (Sept. 2011). “Casein kinase 2 sends extracellular signal-regulated kinase nuclear.” In: *Molecular and cellular biology* 31.17.

Wilson, Maxwell Z et al. (Aug. 2017). “Tracing Information Flow from Erk to Target Gene Induction Reveals Mechanisms of Dynamic and Combinatorial Control”. In: *Molecular cell*.

Worman, Howard J and Gisèle Bonne (June 2007). “"Laminopathies": a wide spectrum of human diseases.” In: *Experimental cell research* 313.10.

Wu, Jianguo et al. (Mar. 2005). “Molecular mechanism of inhibition of survivin transcription by the

GC-rich sequence-selective DNA binding antitumor agent, hedamycin: evidence of survivin downregulation associated with drug sensitivity." In: *The Journal of biological chemistry* 280.10.

Yagil, Gad (1975). "Quantitative Aspects of Protein Induction". In: Elsevier.

Yamamoto, K R and B M Alberts (June 1976). "Steroid Receptors: Elements for Modulation of Eukaryotic Transcription". In: *Annual review of biochemistry* 45.1.

Yan, Chunhong and Paul J Higgins (Jan. 2013). "Drugging the undruggable: transcription therapy for cancer." In: *Biochimica et biophysica acta* 1835.1.

Yang, Edward et al. (Aug. 2003). "Decay rates of human mRNAs: correlation with functional characteristics and sequence attributes." In: *Genome Research* 13.8.

Zeisel, Amit et al. (Sept. 2011). "Coupled pre-mRNA and mRNA dynamics unveil operational strategies underlying transcriptional responses to stimuli." In: *Molecular Systems Biology* 7.

Zhang, H M et al. (Feb. 2008). "Mitogen-induced recruitment of ERK and MSK to SRE promoter complexes by ternary complex factor Elk-1". In: *Nucleic acids research* 36.8.

Zhang, Kai and Bianxiao Cui (Feb. 2015). "Optogenetic control of intracellular signaling pathways". In: *Trends in Biotechnology* 33.2.

Zhang, Kai et al. (Mar. 2014). "Light-Mediated Kinetic Control Reveals the Temporal Effect of the Raf/MEK/ERK Pathway in PC12 Cell Neurite Outgrowth". In: *PloS one* 9.3.

Zhuang, Shougang and Rick G Schnellmann (Dec. 2006). "A death-promoting role for extracellular signal-regulated kinase." In: *Journal of Pharmacology and Experimental Therapeutics* 319.3.

C.2 Acronyms and abbreviations

Full gene names

ΔRAF1:ER	Inducible fusion protein of RAF1 kinase domain and ER hormone binding domain.
AKIRIN2	Akirin-2.
AKT	Family of protein kinases B.
AP-1	Transcription factor complex Ap1.
ARC	Activity-regulated cytoskeleton-associated protein.
ARL5B	ADP-ribosylation factor–like protein 5B.
BAIAP2	Adhesion G protein-coupled receptor B1 (BAI1)–associated protein 2.
BCR-ABL	Fusion protein of breakpoint cluster region gene and abelson murine leukemia viral oncogene homolog 1 gene; oncogenic driver of CML.
BHLHE40	Basic helix-loop-helix family member E40.
BRAF	Rapidly accelerated fibrosarcoma isoform B.
C/EBPδ	CCAAT enhancer binding protein delta.
CD4	Cluster of differentiation 4 receptor.
CDKN1A	Cyclin dependent kinase inhibitor 1A.
CLU	Clusterin.
CRY2	Chryptochrome 2.
CYR61	Cysteine-rich angiogenic inducer 61.
DUSP	Family of dual specificity phosphatases.
EGF	Epidermal growth factor.
EGFR	Epidermal growth factor receptor.
EGR	Family of early growth response proteins.
ELK1	ETS–like tyrosine kinase 1.
ER	Oestrogen receptor.
ERBB2	Erythroblastic oncogene B2 (also named HER2).
ERK	Family of Extracellular signal–regulated kinases.
ERRFI1	ERBB receptor feedback inhibitor 1.
ETS	Family of E26 transformation-specific transcription factors.
ETV5	ETS variant 5.
FADD	Fas-associated protein with death domain.
FasL	Fas receptor ligand.
FGF	Fibroblast growth factor.

Full gene names

FOS	c-Fos, cellular homolog of viral oncogene v-Fos.
FOSB	FOS isoform B.
FOSL1	FOS-like protein 1.
Fra-1	Protein encoded by FOSL1 gene.
GAB1	GRB2-associated-binding protein 1.
GADD45A/B	Growth arrest and DNA-damage-inducible protein alpha/beta.
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase.
GPR50	G protein-coupled receptor 50.
GRB2	Growth factor receptor-bound protein 2.
HDAC	Histone deacetylase.
HER2	Human epidermal growth factor receptor 2 (also named ERBB2).
HGF	Hepatocyte growth factor.
HOMER1	Homer protein homolog 1.
HRG	Heregulin.
hsp90	Heat shock protein 90.
IFNγ	Interferon gamma.
IGF	Insulin like growth factor.
INSIG1	Insulin induced gene 1.
JNK	c-Jun n-terminal kinase.
JUN	c-Jun, cellular homolog of viral oncogene v-Jun.
JUNB	JUN isoform B.
KDM6B	Lysine demethylase 6B.
KLF	Family of krüppel-like factor proteins.
KRAS	Kirsten rat sarcoma.
KRT8	Keratin 8.
MAFF	bZip Maf transcription factor.
MEK	Family of MAPK/ERK kinases.
mTOR	Mechanistic target of rapamycin.
MYC	Cellular homolog of viral myeloblastosis oncogene.
NAGK	N-Acetylglucosamine Kinase.
NELF	Negative elongation factor.
NF-κB	Nuclear factor kappa-light-chain-enhancer of activated B-cells.
NGF	Neuronal growth factor.
PDGF	Platelet-derived growth factor.
PHYB	Phytochrome B.
PI3K	Phosphoinositide 3-kinase.

PKC	Protein kinase C.
PTGS2	Prostaglandin-endoperoxide synthase 2.
QSOX1	Quiescin Sulfhydryl Oxidase 1.
RAB25	Ras-related in brain GTPase 25.
RAF	Family of rapidly accelerated fibrosarcoma kinases.
RAF1	RAF kinase 1.
RalGDS	Ral guanine nucleotide dissociation stimulator.
RAP1	Ras-related protein 1.
RAS	Family of rat sarcoma G-Proteins.
RBM14	RNA Binding Motif Protein 14.
RTKs	Family of receptor tyrosine kinases.
SERPINB9	Serpin family B member 9.
SHC	Src homology 2 domain containing transforming protein 1.
SOS	Son of sevenless.
SPRY	Sprouty homolog.
STAT3	Signal transducer and activator of transcription 3.
TFPI2	Tissue factor pathway inhibitor 2.
TIAM1	T-lymphoma invasion and metastasis-inducing protein 1.
TLR4	Toll-like receptor 4.
TNF	Family of tumour necrosis factors.
TNFAIP3	Tumor necrosis factor alpha-induced protein 3.
TNFRSF12A	Tumor necrosis factor receptor superfamily member 12A.
TP53	Tumor protein p53.
TXNL4B	Thioredoxin-like 4B.
ZBTB38	Zinc finger and BTB domain containing 38.
ZCCHC12	Zinc finger CCHC-type containing 12.
ZFP36	Zinc finger protein 36 homolog.

Other acronyms and abbreviations

[^{n}X]	Notation for an isotope of element X with mass n.
4OHT	4-hydroxy tamoxifen.
4SU	4-thio uridine.
6SG	6-thio guanine.
ActD	Actinomycin D.
API	Application programming interface.
ARE	AU-rich element.
AU-rich	Adenine/Uridine rich.

Other acronyms and abbreviations

AUC	Area under the curve.
BCA	Bicinchoninic acid.
BTB domain	Broad-complex, tramtrack and bric a brac domain.
CBL	Casitas B-lineage lymphoma; E3 ubiquitin-protein ligase.
CCL39	A hamster lung fibroblast cell line.
cDNA	Complementary DNA.
CML	Chronic myeloid leukemia.
CNA	Copy number alteration.
CR1/2/3	Conserved region 1/2/3.
CRD	Cystin-rich domain.
Cre/loxP system	LoxP site-specific Cre recombinase system.
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats; genome editing technique.
CYHX	Cycloheximide.
DEG	Delayed-early gene.
DMEM	Dulbecco's Modified Eagle Medium.
DNA	Deoxyribonucleic acid.
DTA	Dynamic transcriptome analysis.
ELISA	Enzyme-linked immunosorbent assay.
EtOH	Ethanol.
FACS	Fluorescence-associated cell sorting.
FDA	U.S. Food and Drug Administration.
FDR	False discovery rate.
GC-content	Guanine/Cytosine-content.
GEO	Gene Expression Omnibus; online data base for transcriptomic data.
GO	Gene Ontology.
GOF	Gain of function.
GRCh37	Genome Reference Consortium Human Build 37.
GTP	Guanosine triphosphate.
GTPase	Guanosine triphosphate hydrolase.
H1299	A human non-small cell lung cancer cell line.
HEK293	A human embryonic kidney cell line.
HPV	human papillomavirus.
HT22	A mouse hippocampal neuronal cell line.
HUGO	Human Genome Organisation.
ID-miRs	Immediately downregulated micro RNA.

IEG	Immediate-early gene.
ILG	Immediate-late gene.
IQR	Inter quartile range.
K562	A human myelogenous leukemia cell line.
LOF	Loss of function.
LPS	Lipopolysaccharide.
MCF10A	A human mammary gland cell line.
MCF7	A human breast cancer cell line.
mRNA	Messenger RNA.
ncRNA	Non-coding RNA.
NGS	Next generation sequencing.
NSCLC	Non-small cell lung cancer.
ODE	Ordinary differential equation.
PBS	Protein biosynthesis.
PC12	A rat pheochromocytoma cell line.
PFA	Paraformaldehyde.
PGK1	Phosphoglycerate kinase 1.
PMA	Phorbol 12-myristate 13-acetate.
PolII	RNA polymerase II.
PPP1R15A	Protein phosphatase 1 regulatory subunit 15A.
PRG	Primary response gene.
qRT-PCR	Quantitative real-time PCR.
RBD	Ras binding domain.
RBP	RNA-binding protein.
RMA	Robust multi-array average.
RNA	Ribonucleic acid.
RNAi	RNA interference.
rRNA	Ribosomal RNA.
SDDC	Signal duration decoding capacity.
Ser	Serine.
smFISH	Single molecule RNA fluorescence in situ hybridization.
SRE	Serum response element.
SRG	Secondary response gene.
STAR	Bioinformatic read mapping software.
T98G	A human glioblastoma cell line.

Thr	Threonine.
TMM	Trimmed mean of median values; normalisation method for RNA-sequencing experiments.
TPM	Transcripts per million.
tRNA	Transfer RNA.
TSS	Transcription start site.
U0126	MEK inhibitor.
UT	Untreated.
UTR	Untranslated region.
V600E	Valine (V) to glutamate (E) substitution at the 600th position of the BRAF kinase.
wRSS	Sum of weighted squared residuals.
WT	Wild type.

C.3 List of Figures

Danksagung

Danke Nils für die großartige Betreuung meiner Arbeit. Danke, dass ich so enorm viel von Dir und bei Dir lernen durfte. Du hast einerseits unfassbar wertvollen Input gegeben – sowohl auf konzeptioneller Ebene, aber vor allem auch sehr direkte, unmittelbar umsetzbare Ideen und Anregungen gegeben – und andererseits ein Umfeld geschaffen, in dem ich frei lernen, forschen und sein konnte. Ich hätte mir keine bessere Betreuung vorstellen können. Danke Bertram für Dein verlässliches Feedback und Deine unerschütterlich geduldige Art. Danke Anja für Deine unzähligen Antworten auf meine unzähligen Fragen im Labor. Danke Anja, Ela und Nadine für die Erhebung der Microarray- und Protein-Daten. Danke Anja und Emanuel für die Erhebung der 4SU-Daten. Danke Johannes, Mattias und Torsten für die ausschweifenden, stimulierenden und witzigen Diskurse der kleinsten Probleme und größten Ideen. Danke Allen in der Arbeitsgruppe für eine wunderbare, intensive, großartige Zeit. Danke Emanuel und Markus für die fantastische Kollaboration und den reibungslosen, ergiebigen Austausch. Danke an die Berlin School of Integrative Oncology für die Finanzierung meiner Stelle. Danke an das Berlin Institute of Health für die Auszeichnung meiner Arbeit. Danke an Theresa, meine Freunde und meine Familie für alles andere.

Selbständigkeitserklärung

Hiermit erkläre ich, die Dissertation selbstständig und nur unter Verwendung der angegebenen Hilfen und Hilfsmittel angefertigt zu haben. Ich habe mich anderwärts nicht um einen Doktorgrad beworben und besitze keinen entsprechenden Doktorgrad. Ich erkläre, dass ich die Dissertation oder Teile davon nicht bereits bei einer anderen wissenschaftlichen Einrichtung eingereicht habe und dass sie dort weder angenommen noch abgelehnt wurde. Ich erkläre die Kenntnisnahme der dem Verfahren zugrunde liegenden Promotionsordnung der Mathematisch-Naturwissenschaftlichen Fakultät I der Humboldt-Universität zu Berlin vom 27. Juni 2012. Weiterhin erkläre ich, dass keine Zusammenarbeit mit gewerblichen Promotionsberaterinnen/Promotionsberatern stattgefunden hat und dass die Grundsätze der Humboldt-Universität zu Berlin zur Sicherung guter wissenschaftlicher Praxis eingehalten wurden.

Berlin, den 29.11.2018

Florian Sören Uhlitz

DOI: 10.18452/20005